JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 21, No. 2 FEBRUARY, 1950

In This Issue

		Page
Note on Collapsing Cylindrical Shells	THEODORE E. STERNE	73
Steady-State Internal Temperature Rise in Magnet Coil Windings	CLAUDE L. EMMERICH	75
Evaporation and Outgassing in an Inert Atmosphere	R. L. LONGINI	81
Method of Reducing Deflection Defocusing in Cathode-Ray Tubes	R. G. E. HUTTER AND S. W. HARRISON	84
Transmission of Elastic Waves through a Stratified Solid Medium	WILLIAM T. THOMSON	89
Theory of *N* Coupled Parallel Antennas	RONOLD KING	94
Dimensional Analysis of Metal Cutting	D. C. DRUCKER AND H. EKSTEIN	104
Current Build-Up in a Planar Diode	EDWARD H. GAMBLE	108
Thermocouples of the Refractory Metals	F. HAYDN MORGAN AND W. E. DANFORTH	112
Interaction of a Spiral Electron Beam and a Resonant Microwave Cavity	F. K. WILLENBROCK AND S. P. COOKE	114
Geometrical Factors Affecting the Contours of X-Ray Spectrometer Maxima. II. Factors Causing Broadening	LEROY ALEXANDER	126
Determination of Crystallite Size with the X-Ray Spectrometer	LEROY ALEXANDER AND HAROLD P. KLUG	137
Radiofrequency Mass Spectrometer	WILLARD H. BENNETT	143
Strain Induced Grain Boundary Migration in High Purity Aluminum	PAUL A. BECK AND PHILIP R. SPERRY	150
External Field Produced by a Slot in an Infinite Circular Cylinder	SAMUEL SILVER AND WILLIAM K. SAUNDERS	153
Reversal Theorem of Linearized Supersonic Airfoil Theory	M. M. MUNK	159
Impact of a Body on a Water Surface at an Arbitrary Angle	LEON TRILLING	161
Depolymerization by Ultrasonic Irradiation: The Role of Cavitation	ALFRED WEISSLER	171
Three-Stage Electron Microscope with Stereographic Dark Field, and Electron Diffraction Capabilities	M. E. HAINE, R. S. PAGE, AND R. G. GARFITT	173
Letters to the Editor:		
Comments on the Use of Latex Spheres as Size Standards in Electron Microscopy	CHARLES H. GEROULD	183
Erratum: Relationship between Gough-Joule Coefficients and Moduli of Vulcanized Rubbers	F. S. CONANT, G. L. HALL, AND G. R. THURMAN	184
Frictional Loss in Visco-Elastic Substances	B. GROSS	185
Addendum: Interpretation of X-Ray Patterns of Cold-Worked Metal	B. L. AVERBACH AND B. E. WARREN	185
Bibliography on Offshore Petroleum Developments	E. N. KEMLER	185
New Books		186
Index to Advertisers		xxx

BOARD OF EDITORS
ELMER HUTCHISSON, *Editor*

Associate Editors
HOWARD H. AIKEN
OTTO BEECK
JAMES HILLIER
W. JAMES LYONS
PHILIP MORRISON
ERIC REISSNER
RAYMOND J. SEEGER
FREDERICK SEITZ
CHARLES S. SMITH

Publication of the American Institute of Physics

OFFICERS
GEORGE R. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*
JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD
LUIS W. ALVAREZ
J. W. BEAMS
WALLACE R. BRODE
E. U. CONDON
W. F. FAIR, JR.
HARVEY FLETCHER
GEORGE R. HARRISON
W. V. HOUSTON
PAUL KIRKPATRICK
PAUL E. KLOPSTEG
F. W. LOOMIS
PHILIP M. MORSE
J. R. OPPENHEIMER
GEORGE B. PEGRAM
HAROLD K. SCHILLING
JOHN T. TATE
WALLACE WATERFALL
M. W. WHITE

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Assistant Treasurer and Comptroller*
RUTH F. BRYANS, *Publication Manager*
BETTY HOFSTADTER, *Assistant Secretary*
EILEEN B. NEUBERGER, *Circulation Manager*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 free reprints. Instructions will be sent with galley proofs.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices
Complete set: Vol. 1, July, 1931–Vol. 19, 1949—$148.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1950, by the American Institute of Physics. Permission is hereby given to reproduce freely author abstracts in abstract journals.

Journal
of
Applied Physics

Volume 21, Number 2 February, 1950

A Note on Collapsing Cylindrical Shells

THEODORE E. STERNE
Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland
(Received May 23, 1949)

Expressions are obtained for the instantaneous dependence of pressure upon radius within a collapsing cylindrical shell of incompressible liquid when subjected to (a) neither external nor internal pressure or to (b) an external pressure but no internal pressure. Dynamical considerations suggest that very high pressures and radial velocities should be reached.

I.

DYNAMICAL considerations suggest that if an undisturbed cylindrical shell of liquid contracts radially, the velocity of the outer surface should tend to diminish and of the inner surface to increase, and that very high velocities and pressures should be reached. The author thinks that the collapse of a cylindrical liquid shell may bear some relation to details of the behavior of cylindrical and possibly of conical metal shells, caused to collapse by the detonation of external charges of high explosive. The collapse of such shells has already been considered, hydrodynamically, by Birkhoff, MacDougall, Pugh, and Taylor[1] without regard to such details as the instantaneous dependence of pressure upon radial coordinate.

II.

It is felt that the following brief treatment, which has been discussed with Professor Garrett Birkhoff, of an ideal case in which the liquid is incompressible and the collapse radial, without axial motion, may be of sufficient interest to merit a somewhat wider distribution and that it may also perhaps be regarded as a first step toward a detailed theoretical study of the collapse of a metal shell surrounded by an explosive charge.

III.

Consider a long liquid cylindrical shell with external and internal radii r_2 and r_3, respectively, and mass m per unit length. The quantity $r_0^2=r_2^2-r_3^2$ will be conserved as the cylinder contracts if, as is supposed, the motion is radial and the liquid incompressible. Let r_2/r_0 be denoted by x. Then the kinetic energy per unit length is

$$T=(m/2)r_0^2x^2\dot{x}^2\log[x^2/(x^2-1)],$$

and this is conserved if the shell is set contracting radially and left undisturbed. Any element of matter in the shell can be labeled by a Lagrange hydrodynamical coordinate μ defined as the mass, per unit of length, contained between the outside of the shell and a coaxial cylinder through the element. The velocity of any element of matter characterized by a particular value of μ is

$$\frac{\partial}{\partial t}r(\mu,t)=-\left(\frac{2T}{m}\right)^{\frac{1}{2}}\frac{1}{[\log^{\frac{1}{2}}x^2/(x^2-1)][x^2-\mu/m]^{\frac{1}{2}}}.$$

It can be seen from the preceding equation that the velocity of the internal surface approaches infinity as x approaches unity, and that the velocity of every other element of matter in the shell approaches zero as x approaches unity. It follows that during the undisturbed radial contraction of a cylindrical shell, the kinetic energy per unit mass tends to be more and more strongly concentrated toward the center as the contraction proceeds.

IV.

It may readily be shown that the pressure, zero at the outer and inner surfaces, has a maximum at

$$\mu/m=x^2-x^2(x^2-1)\log x^2/(x^2-1).$$

[1] Birkhoff, MacDougall, Pugh, and Taylor, J. App. Phys. **19**, 563 (1948).

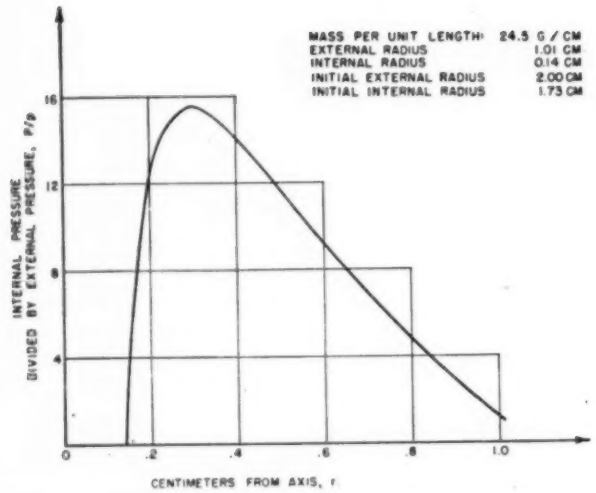


FIG. 1. The distribution of pressure in a radially contracting cylindrical shell of liquid, subjected to pressure over its outer surface.

The maximum pressure occurs at a place about midway between the surfaces when the shell is large, but moves toward the inner surface as the shell contracts. When the external and internal pressures are zero, the pressure is given generally throughout the cylindrical shell by the equation

$$P=\frac{1}{2\pi}\int_0^{\mu}\frac{1}{r}\frac{\partial^2 r}{\partial t^2}d\mu,$$

and has the value

$$P=\frac{T}{\pi r_0^2}\frac{1}{x^2(x^2-1)\log^2 x^2/(x^2-1)}\times\left[\log\frac{x^2}{x^2-\mu/m}-\frac{\mu}{m}\frac{x^2-1}{x^2-\mu/m}\log\frac{x^2}{x^2-1}\right].$$

Enormous pressures can be reached.

V.

If a constant pressure p is exerted on the external surface of the cylindrical shell, initially at rest, and zero pressure on the internal surface, then the resulting velocities and pressures can be calculated. The radial velocity of an element of matter in the shell is given by

$$\frac{\partial}{\partial t}r(\mu,t)=-\left(\frac{2\pi p}{m}\right)^{\frac{1}{2}}r_0\frac{(x_0^2-x^2)^{\frac{1}{2}}}{[\log^{\frac{1}{2}}x^2/(x^2-1)][x^2-\mu/m]^{\frac{1}{2}}},$$

where x_0 is the initial value of x, and the pressure in the shell is given by

$$P=p\left[1-\frac{\log x^2/(x^2-\mu/m)}{\log x^2/(x^2-1)}+\frac{x_0^2-x^2}{x^2(x^2-1)}\frac{\log x^2/(x^2-\mu/m)}{\log^2 x^2/(x^2-1)}-\frac{\mu}{m}\frac{x_0^2-x^2}{x^2(x^2-\mu/m)\log x^2/(x^2-1)}\right]$$

which has a maximum at

$$\frac{\mu}{m}=x^2-\frac{x^2(x^2-1)(x_0^2-x^2)\log x^2/(x^2-1)}{x_0^2-x^2-x^2(x^2-1)\log x^2/(x^2-1)}.$$

VI.

As an example of the application of the preceding general results, the following calculation may be of interest. If $p=10^{11}$ dynes/cm^2, if the initial value of the external radius r_2 is 2 cm, if r_0 is 1 cm, and if the liquid shell has the density of steel so that m is 24.5 g, then when $x=1.01$ the maximum pressure is at a value μ/m of 0.937 and is 15.6×10^{11} dynes/cm^2, which is sixteen times greater than the external pressure. The maximum velocity is then approximately 9.84×10^5 cm per sec., and is of course at the inner surface. In the preceding example the initial value of the internal radius is 1.73 cm and at the stage of contraction considered the external and internal radii are respectively 1.01 cm and 0.14 cm. The pressure is a maximum at 0.29 cm from the axis.

VII.

It seems intuitively plausible that the compressibility of any real liquid would probably cause the velocities and pressures to be less than they would be in an incompressible liquid. The elastic and plastic properties of a metallic shell would involve further departures from the ideal case of an incompressible liquid. The author is now unable to predict the magnitude of such departures. The magnitude may be limited, however, in cases of axially progressive collapse of the type considered by Birkhoff, MacDougall, Pugh, and Taylor.[1]

Steady-State Internal Temperature Rise in Magnet Coil Windings

CLAUDE L. EMMERICH*
Applied Science Research Laboratory, University of Cincinnati, Cincinnati, Ohio
(Received May 16, 1949)

The temperature distribution in an electric coil winding is determined by solving the differential equations which apply to the system approximately. The maximum and average temperature rises are given for several types of windings. A distinction is made between small and large temperature rise; the treatment of the former neglects local changes in resistivity and thermal conductivity caused by the increase in temperature. A correction factor is computed to obtain the actual temperature rise from the much simpler calculations for the maximum temperature rise which would occur if the resistance were independent of temperature. It is found that non-uniform windings, wound of several different wire sizes to minimize the resistance, exhibit the same temperature rise characteristics as conventional uniform windings of zero curvature.

INTRODUCTION

THE differential equation which applies to the steady-state temperature distribution in a magnet coil winding is Poisson's equation:

$$\nabla^2 T = -Q/K, \tag{1}$$

where T is the temperature at any point in the winding, Q is the rate with which heat is generated per unit volume per unit time, and K is the heat conducted per unit area per unit time per unit temperature gradient. The symbol ∇^2 denotes the Laplacian differential operator. In Cartesian coordinates, Eq. (1) would read

$$\frac{\partial^2 T}{\partial x^2}+\frac{\partial^2 T}{\partial y^2}+\frac{\partial^2 T}{\partial z^2}=-\frac{Q}{K}.$$

A coil winding is a heterogeneous structure, including metallic wire, surrounded by an insulating coating, and separated by insulating spaces. The term "insulating" refers qualitatively to both thermal and electrical properties. Thus, Eq. (1) as applied to magnet coils must represent a very complex system. However, for practical applications, it is sufficiently accurate to assume that the winding is a homogeneous body, with definite values of Q and K, averaged over the local discontinuities.

For a winding wound of one wire size, the value of Q is determined from the equation

$$Q=\rho A^2/(fL^2h^2), \tag{2}$$

where ρ = resistivity of the metal in the wire, ohm-inch; A = number of ampere-turns; f = space factor (cross-section area of metal divided by area of the winding Lh); L = length of winding, inch; h = thickness of winding, inch, $=b-a$ (see Fig. 1).

The value of K is not easy to predict from the physical constants of the coil. Moore[1] has made a theoretical investigation related to the problem of determining K; however, it appears that the theory has not yet been developed to give values of K more easily and accurately than experimental methods. Since K is an average of the thermal conductivities of the materials in the coil, it will be constant or variable according to the thermal characteristics of the wire and of the insulating materials of which the coil is composed. If the internal temperature rise in the winding is small, K may be assumed constant; and Q will also be nearly constant because the resistivity of the wire is not changed very much by a small change in the temperature.

Thus the solution of Eq. (1) is simplified considerably in the range of small temperature rises.

SMALL TEMPERATURE RISE

In particular, the small temperature rise conditions will be applied to a winding in the shape of a hollow cylinder (Fig. 1). If the winding is thin in comparison with its inside and outside diameters and its length, the curvature of the turns may be neglected, and the flow of heat may be assumed in the radial direction only, satisfying approximately the ordinary differential equation

$$(d^2T')/(dx^2)=-Q/K, \tag{3}$$

where x is the radial distance from the center line of the winding to the point whose temperature is T'. This equation will apply also approximately to a winding on a rectangular core.

Two sets of boundary conditions will be adopted as representing actual operating conditions of long magnet coil windings: (a) Core thermally insulated, outside surface at a fixed temperature T_s and (b) both core surface and outside surface at the same fixed temperature T_s.

These boundary conditions have been chosen in the interest of simplicity. It is evident that an actual coil does not operate under such conditions; however, to obtain a first approximation, some ideal cases should be considered. The deviations for actual coils may be judged from work reported in the literature.[2-4]

* Present address: 270 Harvard Street, Cambridge, Massachusetts.

[1] A. D. Moore, *Fundamentals of Electrical Design* (McGraw-Hill Book Company, Inc., New York, 1927), p. 95.

[2] G. L. Moses, Prod. Eng. **10**, 317 (1939).

[3] H. C. Stewart and L. C. Whitman, Trans. A.I.E.E. **63**, 763 (1944).

[4] H. M. Beede and B. M. Cain, Elec. Eng. **66**, 87 (1947).

Using boundary conditions (a) in the solution of Eq. (3), the temperature distribution in the winding is found to be

$$\Theta'(x)=T'(x)-T_s=(Q/2K)(b^2+2ax-x^2-2ab), \quad (4)$$

where Θ' is defined as the temperature rise above the surface temperature T_s. The maximum temperature rise in the winding will occur at the inside surface ($x=a$), where

$$\Theta_m'=\rho A^2/(2fKL^2). \quad (5)$$

Equation (5) is plotted in terms of the dimensionless variable $2fKL^2\Theta'/\rho A^2=1$ in Fig. 2.

A useful criterion of internal temperature rise is the average temperature rise, obtained from a measurement of the increase in total resistance of the winding. The average temperature is defined as the temperature to which the entire coil would have to be raised in order

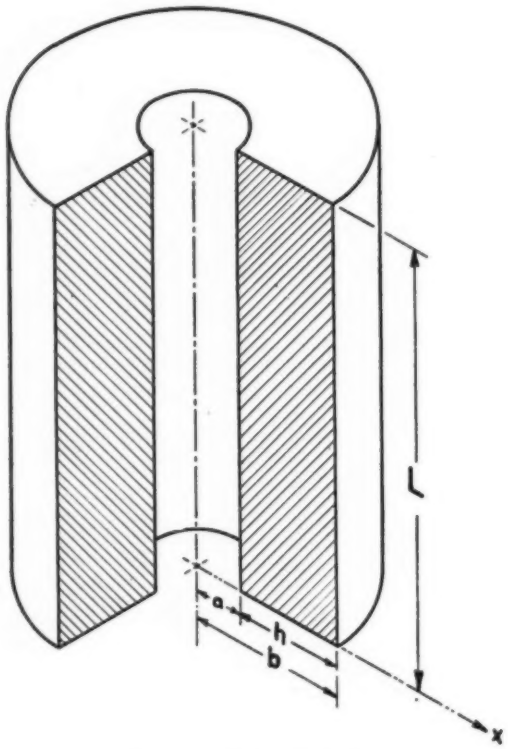


FIG. 1. Cylindrical winding.

to have the same increase in resistance as it actually has in the temperature field created by the heat generated in the winding. Expressed analytically, the average temperature rise is given by the equation

$$\Theta_{hv}=\left(\int\Theta dR\right)\Big/\left(\int dR\right), \quad (6)$$

where dR is the resistance of an elemental section of the winding whose temperature rise is Θ.

The temperature distribution as a function of the radius given by Eq. (4), together with the known distribution of resistance as a function of x, permits evaluation of the average temperature rise; which is found to be, for a winding with boundary conditions (a),

$$\Theta_{hv}'=\rho A^2/(3fKL^2). \quad (7)$$

This is plotted in Fig. 2 as $2fKL^2\Theta'/\rho A^2=\frac{2}{3}$.

For practical purposes it should be noted that the average and maximum temperature rises are in the ratio 2:3 for the thin cylindrical winding and small temperature rise.[5]

If boundary conditions (b) are applied to Eq. (3), the temperature distribution takes the form

$$\Theta'(x)=(Q/2K)(ax+bx-x^2-ab). \quad (8)$$

The maximum temperature rise will be the temperature rise at the center of the coil, where

$$\Theta_m'=\rho A^2/(8fKL^2). \quad (9)$$

In Fig. 2 this is plotted as $2fKL^2\Theta'/\rho A^2=\frac{1}{4}$.

The average temperature rise of the winding with boundary conditions (b) is

$$\Theta_{hv}'=\rho A^2/(12fKL^2) \quad (10)$$

plotted as $2fKL^2\Theta'/\rho A^2=\frac{1}{6}$ in Fig. 2.

For a thick, long, cylindrical winding, the assumption of negligible curvature, which led to the simple Eq. (3), cannot be applied. For this case, Poisson's Eq. (1) reduces to the ordinary differential equation

$$\frac{d^2T'}{dx^2}+\frac{1}{x}\frac{dT'}{dx}=-\frac{Q}{K}, \quad (11)$$

where x again denotes the radius of the point of temperature T'.

With boundary conditions (a), the temperature rise as a function of the radius is given by the expression

$$\Theta'(x)=(Q/4K)[b^2-x^2-2a^2\log_e(b/x)]. \quad (12)$$

The maximum temperature rise will occur at the inside of the winding, where

$$\Theta_m'=\frac{\rho A^2}{2fKL^2}\cdot\frac{r^2-1-2\log_e r}{2(r-1)^2}. \quad (13)$$

In this expression, $r=b/a$ is the thickness ratio of the winding, defined as the fraction outside radius divided by inside radius. Equation (13) will be found to represent actual conditions more accurately than the equation given for this case by Moore.†

Equation (13) may be represented by the following

[5] M. Jakob, Trans. A.S.M.E. **65**, 593 (1943). Also M. Jakob and G. A. Hawkins, *Elements of Heat Transfer and Insulation* (John Wiley and Sons, Inc., New York, 1942), Chapter V.

† See reference 1, p. 98.

series expansion

$$\Theta_m' = \frac{\rho A^2}{2fKL^2}\left[\frac{(r+1)^2}{2(r^2+1)} - \frac{1}{3}(r-1)\left(\frac{r+1}{r^2+1}\right)^3 - \frac{1}{5}(r-1)^3\left(\frac{r+1}{r^2+1}\right)^5 - \cdots\right]. \quad (14)$$

The relation between $2fKL^2\Theta'/\rho A^2$ and r, as given by Eq. (13) is shown in Fig. 2 by the curve marked "Θ_m' uniform (a)." It is evident from this curve, as well as from the expansion (14), that for a thickness ratio of one, the special case of the thin winding is obtained. The other extreme, representing a winding in the shape of a solid cylinder, is seen to be approached as the thickness ratio becomes greater. For $r \to \infty$, Eqs. (13) and (14) reduce to the value

$$\Theta_m' = \tfrac{1}{2}\rho A^2/(2fKL^2). \quad (15)$$

which is just one-half of the expression for the maximum temperature rise of the thin winding [Eq. (5)]. It is indicated at the right edge of Fig. 2.

The average rise for the thick hollow cylinder with boundary conditions (a) can be determined from the temperature distribution $\Theta'(x)$ [Eq. (12)], and from the fact that the resistance of an elemental section of the hollow cylinder is proportional to the expression xdx. The average temperature rise is thus found to be

$$\Theta_{Av}' = \frac{\rho A^2}{2fKL^2} \cdot \frac{r^2 - 3 + [4/(r^2-1)]\log_e r}{4(r-1)^2}. \quad (16)$$

Figure 2 shows this curve, marked "Θ_{Av}' uniform (a)."

The expansion of Eq. (16) is given by the following series

$$\Theta_{Av}' = \frac{\rho A^2}{2fKL^2} \cdot \frac{(r+1)^2}{(r^2+1)}\left[\frac{1}{4} + \frac{1}{3(r^2+1)^2} + \frac{(r^2-1)^2}{5(r^2+1)^4} + \cdots\right]. \quad (17)$$

If boundary conditions (b) are applied to the thick, long, hollow cylindrical winding [Eq. (11)], the temperature rise distribution will be found to be

$$\Theta'(x) = \frac{Q}{4K}\left(b^2 - x^2 - \frac{b^2-a^2}{\log_e(b/a)}\log_e\frac{b}{x}\right). \quad (18)$$

The radius x_m of the surface at which the maximum temperature rise occurs is found from the equation

$$x_m^2 = (b^2-a^2)/[2\log_e(b/a)].$$

The temperature rise at that surface will be

$$\Theta_m' = \frac{\rho A^2}{8fKL^2}\left\{\frac{2r^2}{(r-1)^2} - \frac{r+1}{(r-1)\log_e r}\left[1 + \log_e\frac{2r^2\log_e r}{r^2-1}\right]\right\} \quad (19)$$

which is plotted in Fig. 2 as the curve marked "Θ_m' uniform (b)."

The average temperature rise with boundary conditions (b) is given by the equation

$$\Theta_{Av}' = \frac{\rho A^2}{2fKL^2} \cdot \frac{r^2 + 1 - [(r^2-1)/\log_e r]}{4(r-1)^2} \quad (20)$$

shown in Fig. 2 by the curved marked "Θ_{Av}' uniform (b)."

In a non-uniform winding designed for minimum resistance,[6] the wire diameter varies as the square root

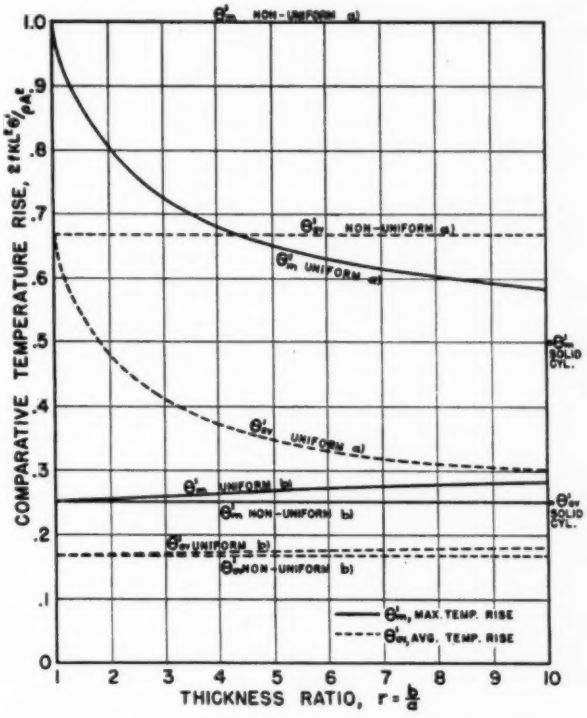


FIG. 2. Maximum and average temperature rise with constant resistivity.

of the radius of the turns in the winding. Thus the value of Q is given by the formula

$$Q = \rho A^2/[fL^2x^2(\log_e r)^2]. \quad (21)$$

Assuming small temperature rise conditions, both ρ and K are constant; hence the temperature field in a non-uniform cylindrical winding is represented by the equation

$$\frac{d^2T'}{dx^2} + \frac{1}{x}\frac{dT'}{dx} + \frac{1}{x^2}\cdot\frac{\rho A^2}{fKL^2(\log_e r)^2} = 0. \quad (22)$$

With boundary conditions (a), the temperature distribution is found to be

$$\Theta'(x) = \frac{\rho A^2}{2fKL^2(\log_e r)^2}\log_e\frac{b}{x}\log_e\frac{bx}{a^2} \quad (23)$$

[6] C. L. Emmerich, Am. Math. Monthly **55**, 613 (1948).

which yields as the maximum temperature rise the value already found in Eq. (5). However, while Eq. (5) was found as the solution to the problem of the uniform winding of thickness ratio one only, the solution just determined is independent of the thickness ratio, hence representable by the straight line marked "Θ_m' non-uniform (a)" in Fig. 2.

Experimental data[7] indicate that the practical approximations to ideal non-uniform windings[8] exhibit the same temperature rise characteristics as the ideal windings. Thus the solution just determined for the ideal non-uniform windings is valid for practical work also.

The remarkable property that non-uniform windings exhibit the same temperature rise, independent of thickness ratio, is shown also by the solution with boundary conditions (b), as indicated in Fig. 2 by the straight line "Θ_m' non-uniform (b)," as well as the average temperature rises "Θ_{A_t}' non-uniform (a)" and "Θ_{A_t}' non-uniform (b)."

The preceding results show quantitatively to what extent overheating may be expected in multiple-wire-size windings, when thin wire is wound near the core as suggested by Emmerich.[6] The conclusion is that the temperature rise is the same as for a uniform winding of the same number of ampere-turns, but wound on a core with zero curvature.

LARGE TEMPERATURE RISE

Insulating materials which were developed recently have made increased operating temperatures possible for magnet coil windings.[9] In order to take full advantage of these new materials, it is imperative to have correct design data at hand. These data are obtainable in part from the solution of Poisson's Eq. (1).

When large temperature rises are considered, it is no longer possible to assume that Q and K are constant. The rate of heating, Q, is a linear function of the temperature, because it contains as a factor the resistivity ρ, which is known to vary linearly with temperature in the range normally expected for operating conditions of electrical equipment. The average thermal conductivity K, being a function of the conductivities of the materials in the coil, may be expected to vary with temperature also; however, it appears difficult to predict the relationship between K and T for any given winding. Experimentally, a coil of Formex wire, immersed in kerosene, was found to have a constant thermal conductivity of 0.020 watt/°C-inch over a temperature range from 12°C to 65°C; while the same coil in air exhibited an increase in K from 0.0057 to 0.0072 watt/°C-inch as the temperature changed from 33°C to 84°C.[7]

Thus, at least for a Formex winding immersed in kerosene, the right side of Eq. (1) should be expressed as a linear function of T. It will be helpful in applying the boundary conditions to let the temperature be represented by the sum of the surface temperature and temperature rise:

$$T=T_s+\Theta.$$

The differential operator ∇^2 will then apply directly to the temperature rise Θ, while the linear function in T is replaced by a linear function in Θ. Poisson's equation then reads as follows

$$\nabla^2\Theta=-u^2\Theta-w^2 \tag{24}$$

where

$$u^2=\frac{\alpha_0\rho_0A^2}{fKL^2h^2} \quad \text{and} \quad w^2=\frac{\rho_0A^2}{fKL^2h^2}(1+\alpha_0T_s).$$

In these expressions, ρ_0 is the resistivity of the metal in the wire at 0°C, and α_0 is the temperature coefficient of resistivity of the wire at 0°C. The boundary conditions are simplified to the extent that at the outside surface of the winding, the temperature rise is zero.

Jakob[5] has published the solutions of Eq. (24) for the solid cylinder ($r\to\infty$) and the flat plate ($r=1$). The maximum temperature rise in the case of the flat winding ($r=1$) is given as follows, in the notation of this paper,

$$\Theta_m=\left(T_s+\frac{1}{\alpha_0}\right)\left[\sec\frac{A(\alpha_0\rho_0)^{\frac{1}{2}}}{L(fK)^{\frac{1}{2}}}-1\right]. \tag{25}$$

From the expansion of the secant function it is readily found that Eq. (25) is approximated by the series

$$\Theta_m=\left(T_s+\frac{1}{\alpha_0}\right)\left[\frac{1}{2!}u^2h^2+\frac{5}{4!}(u^2h^2)^2+\frac{61}{6!}(u^2h^2)^3+\cdots\right]. \tag{26}$$

It is of course possible to use Eq. (25), or its expansion (26), directly in making temperature rise calculations. However, for the purpose of comparison with the design data obtained from Eq. (5), which neglects changes in resistivity, a correction factor Θ_m/Θ_m' has been computed.

If the value of ρ in Eq. (5) is taken as the resistivity of the wire at temperature 0°C, it will be seen that the temperature rise computed from Eq. (5) is

$$\Theta_{m0}'=u^2h^2/2\alpha_0. \tag{27}$$

Substituting this expression in Eqs. (25) and (26) leads to the functional equation

$$\Theta_m=\left(T_s+\frac{1}{\alpha_0}\right)[\sec(2\alpha_0\Theta_{m0}')^{\frac{1}{2}}-1] \tag{28}$$

[7] C. L. Emmerich, thesis, University of Cincinnati (1949), Section V.
[8] C. L. Emmerich, Elec. Eng. **68**, 185 (1949).
[9] G. L. Moses and J. J. Torok, Trans. A.I.E.E. **65**, 412 (1946).

and to the series expansion

$$\theta_m=\left(T_s+\frac{1}{\alpha_0}\right)\left[\alpha_0\theta_{m0}'+\frac{5}{4!}(2\alpha_0\theta_{m0}')^2+\frac{61}{6!}(2\alpha_0\theta_{m0}')^3+\cdots\right]. \quad (29)$$

If the value of ρ in Eq. (5) is taken as the resistivity of the wire at temperature T_s, that is, if

$$\rho_s=\rho_0(1+\alpha_0 T_s)$$

as is often done in coil design, then substitution yields the functional equation

$$\theta_m=\left(T_s+\frac{1}{\alpha_0}\right)\left[\sec\left(\frac{2\theta_{ms}'}{T_s+1/\alpha_0}\right)^{\frac{1}{2}}-1\right] \quad (30)$$

and the corresponding expansion

$$\theta_m=\theta_{ms}'+\frac{5}{6}\cdot\frac{1}{(T_s+1/\alpha_0)}(\theta_{ms}')^2+\frac{61}{90}\cdot\frac{1}{(T_s+1/\alpha_0)^2}(\theta_{ms}')^3+\cdots. \quad (31)$$

Dividing Eq. (30) or (31) through by θ_{ms}' yields the correction factor θ_m/θ_{ms}', which is a function of θ_{ms}' with T_s as a parameter. This family of curves is plotted in Fig. 3, with $1/\alpha_0=234.5$°C.

An example will illustrate the use of these curves: At a surface temperature $T_s=40$°C, the uncorrected temperature rise obtained from Eq. (5) for a particular winding is 84°C, using the value of ρ_s corresponding to $T_s=40$°C. On the curve marked $T_s=40$°C in Fig. 3, the correction factor corresponding to 84°C is 1.35. Hence, the true temperature rise is $1.35\times84=113$°C. The "hot spot" in the coil will therefore reach a temperature of $40+113=153$°C.

It should be kept in mind in using these data that the boundary conditions are idealized. The assumption is made that no heat escapes through the core or through the ends of the coil, and that the temperature at the surface is constant. In actual windings it is probable that heat does escape through the core and ends, thus making the foregoing calculations fall on the safe side; however, the temperature at the surface of the coil may be quite far from the ambient temperature, and it may vary along the surface. This tends to make the calculations uncertain. To take into account the end leakage leads to rather involved computations, as shown by Higgins.[10] Actual surface temperatures are probably obtained easiest by experimental methods.[2–4]

It should be emphasized that Eq. (30) and Fig. 3 apply only to windings of thickness ratio one. Since actual windings are usually in the shape of hollow cylinders, the following computations have been made to evaluate the corrections necessary to allow for the curvature of the winding. It will be found, as in the case of small temperature rise, that the winding of thickness ratio one has a higher maximum temperature rise than a winding in the shape of a hollow cylinder.

For a long, thick, cylindrical winding (heat flow in the radial direction only), with boundary conditions (a), the solution of the equation

$$\frac{d^2\theta}{dx^2}+\frac{1}{x}\frac{d\theta}{dx}+u^2\theta=-w^2 \quad (32)$$

will permit calculation of the temperature rise distribution

$$\theta(x)=\frac{w^2}{u^2}\left[\frac{J_1(au)Y_0(xu)-J_0(xu)Y_1(au)}{J_1(au)Y_0(bu)-J_0(bu)Y_1(au)}-1\right] \quad (33)$$

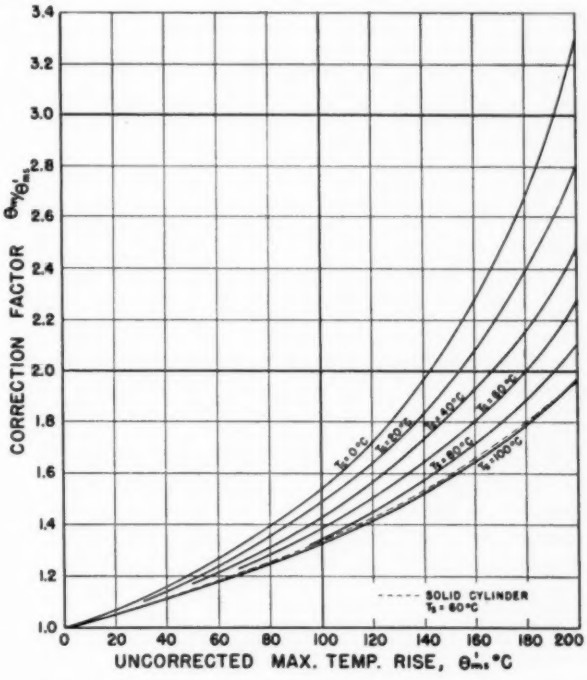


FIG. 3. Correction factor for maximum temperature rise in thin windings.

and of the maximum temperature rise

$$\theta_m=\left(T_s+\frac{1}{\alpha_0}\right)\times\left\{\frac{2}{\pi au[J_1(au)Y_0(bu)-J_0(bu)Y_1(au)]}-1\right\}. \quad (34)$$

J and Y are Bessel functions of the first and second kind, their order being indicated by subscripts.

[10] T. J. Higgins, Trans. A.I.E.E. **64**, 190 (1945) and Trans. A.S.M.E. **66**, 665 (1944).

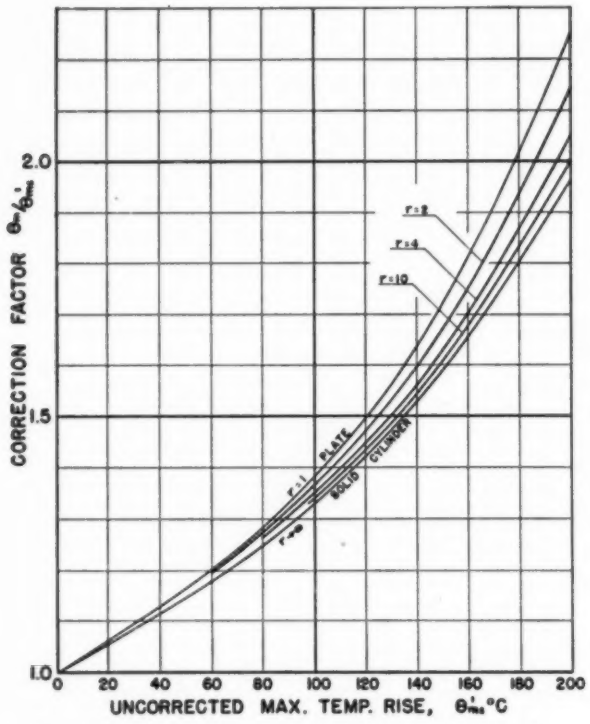


FIG. 4. Correction factor for maximum temperature rise in thick windings, $T_s = 60°C$.

Again, Eq. (34) might be used directly in making temperature rise calculations. However, for the purpose of comparison with design data obtainable from Eq. (14), a correction factor Θ_m/Θ_{ms}' has been computed and plotted in Fig. 4, for the surface temperature $T_s = 60°C$, and using the value of the reciprocal of the thermal coefficient of resistance $1/\alpha_0 = 234.5°C$.

This family of curves is intended to demonstrate that the difference in correction factors for windings of various thickness ratios, other conditions being the same, is rather small. Few actual windings exceed the thickness ratio $r = 2$; thus the approximation made by using the curves in Fig. 3 is not serious. For the purpose of comparison, the curve for a solid cylindrical winding at $T_s = 60°C$ has been shown dotted in Fig. 3.

It may be stated here for the sake of completeness that the non-uniform winding designed for minimum resistance at small temperature rise yields exactly the same temperature rise conditions as the curves in Fig. 3, independent of its thickness ratio.

Up to this point, Poisson's Eq. (1) has been solved for cylindrical windings with Q and K constant; also with K constant and Q a linear function $Q(T)$. If it is assumed that K is also a linear function $K(T)$, then the differential equations become much more complicated, making a solution in closed form difficult. The only exception occurs when $K(T)$ is a multiple of $Q(T)$; that is, when the linear function $K(T)$ would be zero at the same temperature as the temperature for which $Q(T)$ would be zero. Then the right side of Eq. (1) would reduce to a constant again. In practical terms, this requires that the thermal conductivity of the winding must be proportional to the resistivity of the metal in the wire at all temperatures. Ordinary coil windings are not expected to have this special property.

If the function $1/K$ is linear in T, then the right side of Eq. (1), being the product of two linear functions, would become a quadratic function in T. This case has already been considered by Davies,[11] who assumed K to be constant, and $Q(T)$ a quadratic function. The solution in this case contains elliptic functions. It appears that the complete and detailed results of this case are not very useful until it has been established that the thermal conductivities of actual windings are expressed correctly by linear functions $1/K$.

ACKNOWLEDGMENT

Grateful acknowledgment is made to the Stephen H. Wilder Foundation, and to the staff of the Applied Science Research Laboratory, University of Cincinnati, for their cooperation on this project.

[11] E. S. Davies, Trans. A.S.M.E. 65, 602 (1943).

Evaporation and Outgassing in an Inert Atmosphere

R. L. LONGINI
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received June 9, 1949)

Experiments on the evaporation outgassing of antimony in vacuum and in an inert atmosphere show that an inert atmosphere is advantageous for outgassing antimony if loss of antimony by evaporation is intollerable. A simple theory based on well-established kenetic theory considerations is given which may explain and did predict the results.

INTRODUCTION

THE introduction of an inert gas into a vacuum system changes the conditions limiting the rates of both evaporation and outgassing. In a vacuum these rates depend on reflection or accommodation coefficients, whereas in the presence of an inert atmosphere diffusion is the rate-determining process.

It has been known for some time that an atmosphere reduces the rate of evaporation of tungsten in lamps or of magnesium[1] in an inert atmosphere.

In the present study, the actual rates were determined experimentally by use of a microbalance on which loss of sorbed gasses and of the parent body of antimony were found by direct weighing of test samples. A helium atmosphere was used. Results indicate that the introduction of an inert atmosphere may prove useful in outgassing volatile materials.

THEORY

The theory is rather crude and may neglect some important effects. It is important to note that the theory came first and predicted the observed results. It is presented as it may prove useful in some other cases.[2]

Let us assume that the test sample is in equilibrium with its various vapors. The vapor of the parent material (antimony in the experiments) is in equilibrium with the solid body, and other vapors are in equilibrium with their sorbed phases.

Suppose we examine any one of the vapors closely. Let n_{ie} be the number of molecules per unit volume of the ith vapor in the space at equilibrium. Now kinetic theory shows us that $n_{ie}\bar{c}_i/4$ of these will cross a unit area per second where $\bar{c}_i$ is the mean velocity of the ith vapor.[3] This will also be the number of molecules striking each unit area of the condensed material, and a fraction a_i will stick. The fraction a_i is known as the accommodation coefficient. At equilibrium, the rate of condensation or sticking must equal the gross rate of evaporation of the ith component R_i, which gives

$$R_i=n_{ie}\bar{c}_i a_i/4. \tag{1}$$

(If a good vacuum is maintained the net rate of evaporation must be given by R_i above, as the vapor will be removed and no condensation from a vapor phase will take place.)

Suppose we admit an inert atmosphere. The evaporating vapors will now have to diffuse away from the condensed phase. Let R_{pi} be this evaporation rate when the inert gas pressure is p. Now $R_{pi}=D_{pi}(\text{grad}n_i)_0$ where D_{pi} is the diffusion coefficient of the ith vapor through the inert atmosphere of pressure p and n_i is the number of molecules of the ith vapor per cubic centimeter. The gradient is taken at the condensing surface as indicated by the subscript zero.

Now, from our previous discussion, we conclude that the vacuum or gross rate R_i is equal to the net rate R_{pi} plus the rate of recondensing $n_{i0}a_i\bar{c}_i/4$, where n_{i0} is the molecular density at the surface.

We can state

$$\frac{R_{pi}}{R_i}=\frac{D_{pi}(\text{grad}n_i)_0}{n_{i0}\bar{c}_i a_i/4+D_{pi}(\text{grad}n_i)_0}. \tag{2}$$

From kinetic theory we have $D_{pi}=g_iT^{\frac{1}{2}}p^{-1}$ and $\bar{c}_i=h_iT^{\frac{1}{2}}$, where g_i and h_i are proportionality constants. Putting these in the above and dividing numerator and denominator each by $g_i(\text{grad}n_i)_0T^{\frac{1}{2}}$ we have

$$\frac{R_{pi}}{R_i}=\frac{T/p}{T/p+a_in_{ip}h_i/4g_i(\text{grad}n_i)_0}. \tag{3}$$

In the steady state it can be shown that $n_{i0}/(\text{grad}n_i)_0$ is independent of n_i and is a function of geometry only. (In the one-dimensional case, for example, $n_{i0}/(\text{grad}n_i)_0$ equals the distance between the evaporating surface and the remote absorbing surface.) It is therefore essentially the same for all components. We can rewrite the above as

$$R_{pi}=R_i\frac{T/p}{a_i/G_i+T/p}, \tag{4}$$

where G_i takes into account both the geometry and the proportionality constants for the ith vapor. Now the diffusion coefficients for the various vapors are of the same order of magnitude so that the G_i and thus the ratios R_{pi}/R_i of various components have large differences only because of large differences in the accommodation coefficients.

[1] E. A. Gulbransen, Trans. Electrochem. Soc. 87, 597 (1945).
[2] For comparison see I. Langmuir, Phys. Rev. 2, 329 (1913).
[3] For example, *Cobine, Gaseous Conductors* (McGraw-Hill Book Company, Inc., New York, 1941), p. 19.

The accommodation coefficient for the parent material will probably be close to unity, whereas that for various sorbed vapors may be very small. Thus a pressure of inert gas may do much more to reduce the net evaporation for the parent material than for the sorbed components. The ratio of rates of evaporation of two components is thus pressure dependent

$$\frac{R_{p1}}{R_{p2}}=\frac{R_1}{R_2}\frac{T/p+a_2/G_2}{T/p+a_1/G_1}. \tag{5}$$

Anticipating our experimental results with plated antimony and its various sorbed vapors,

$$\frac{R_{p(Sb)}}{R_{p2}}=\frac{R_{(Sb)}}{R_2}\frac{T/p+3.3\times10^4}{T/p+4.1\times10^6} \tag{6}$$

where the subscript *Sb* refers to plated antimony, the subscript 2 refers to an effective "mean" of sorbed vapors. The pressure is in atmospheres, the temperature in degrees Kelvin. In this case the value of $n_{i0}/(\text{grad}n_i)_0$ was about 25 cm. It is apparent that a moderate pressure will make the T/p value small and that the rate of evaporation $R_{p(Sb)}$ will be reduced far more than the rate of outgassing R_{p2}.

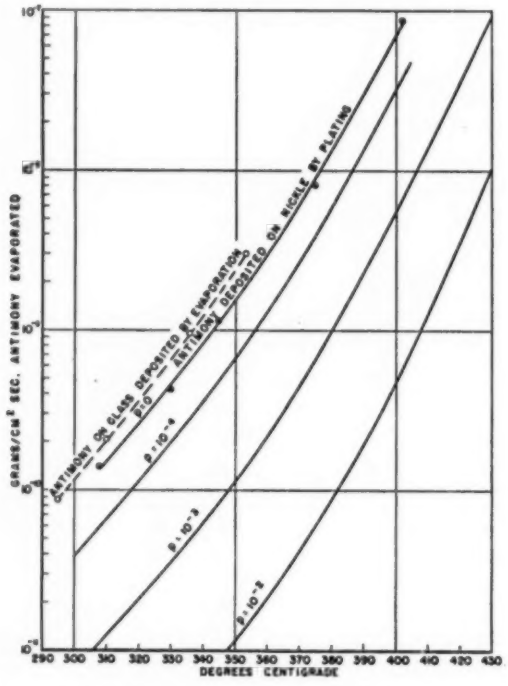

GRAMS/CM² SEC. ANTIMONY EVAPORATED
DEGREES CENTIGRADE
290 300 310 320 330 340 350 360 370 380 390 400 410 420 430
ANTIMONY ON GLASS DEPOSITED BY EVAPORATION
ANTIMONY DEPOSITED ON NICKEL BY PLATING
p=10⁻⁴
p=10⁻³
p=10⁻²

FIG. 1. The rate of evaporation of plated antimony. The rate of evaporation of plated antimony is shown as a function of temperature and helium pressure. The pressures of the isobaric curves are expressed in atmospheres. The vacuum rate of evaporation of antimony from glass is also shown. The experimental points for the vacuum rates are given but those for inert atmospheres were very scattered, varying by a factor of two from the curves shown. Since this is influenced by geometry in any case, the curves are given only to show the order of magnitude of the effect.

EXPERIMENTAL PROCEDURE AND RESULTS

The experiments to determine evaporation rates were carried out on Dr. E. A. Gulbransen's microbalance.[4] This instrument is enclosed in glass and can be evacuated or filled to pressures up to atmospheric with nearly any gas. The sample is suspended from one side of the balance by a tungsten wire so positioned that the sample hangs in a tube which can be heated by a furnace. The counterweight has the same density as the sample to reduce buoyancy effects when gas is admitted.

The balance is calibrated in micrograms and reads the degree of unbalance between the constant counterweight and the sample under investigation. The readings are made with a vernier microscope and are reproducible under most conditions to within one-half a microgram.

Two types of samples were used. The first consisted of 0.005-in. steel shims 1×6 centimeters in extent, plated on both sides with a total of 16 milligrams of antimony from an antimony chloride plating bath. This type was used to determine the rate of evaporation of outgassed plated antimony, the rate of outgassing, and the effect of helium on these rates. The second type consisted of microscope cover glasses each 11×50 millimeters by 0.25 millimeter thick. These were coated with antimony which was deposited by evaporating it from a tungsten wire on which it had been plated. The resultant coating had a surface density of about 16 μg cm^{-2} on each side. These slides were used for determining the rate of evaporation from an evaporated antimony film on glass.

For determining the rate of evaporation of antimony the samples were first outgassed. When the rate of weight loss was constant with time at any one temperature, the sample was believed to be outgassed. This was cross-checked by raising the temperature for a while, then bringing it back to the original temperature. A return of the rate of weight loss to its original steady value was always found.

The various rates were determined by plotting the loss of weight against time at a fixed temperature and observing the slope. During outgassing the resulting plots were curves with steadily, but slowly, decreasing slopes. The slopes of the steady state lines corresponding to outgassed samples were found and, after being put in terms of grams cm^{-2} sec.$^{-1}$, are plotted against temperature in Fig. 1. The rates of evaporation of both forms of antimony have the same slope when plotted against temperature but the evaporated film has a 40 to 50 percent higher rate. This is indicative of the same heat of vaporization but probably of a higher accommodation coefficient for the evaporated film. This indicates an accommodation coefficient of less than 0.7 for plated antimony, a very low value compared to the assumed value of unity used in calculating the calibra-

[4] E. A. Gulbransen, Rev. Sci. Inst. **15**, 201 (1944).

tion of the Knudsen gauge. The apparent higher accommodation coefficient for the evaporated film is probably due to a granular structure similar to those seen with an electron microscope for most metal films. The condensing antimony molecule, falling deep into the layer, might have to undergo several reflections before emerging.

It is interesting to note that the evaporation rate of antimony from the glass slide was independent of the amount of antimony on the slide until this was under 2 $\mu g\ cm^{-2}$. Below this the rate dropped slowly to one-half its former value, then dropped suddenly to zero at which point all the antimony had evaporated.

Outgassed samples were observed when a helium atmosphere was introduced. It was found that even slight impurities in the helium had a very decided effect on the results. The helium was passed through hot, partially oxidized copper and then through a drying tube. This should have removed hydrogen, oxygen, and water vapor. Even after great care with this process, the rates determine with helium present were not reproducible with precision. The determined effect of pressure may be off by 50 percent. The figure shows how the rate of evaporation is dependent on temperature and pressure.

Determining the effect of an inert gas pressure on the rate of outgassing was much more difficult. First of all, the outgassing rate in vacuum is not constant so that the pressure effect had to be measured rapidly. Secondly, the effect was so small that 0.1 atmosphere of helium was needed to measure it. These conditions required a much more rapid introduction of gas than the previously described purifier could handle. A long total obstruction activated charcoal trap was built to purify the gas more rapidly.

The tests were carried out as quickly as possible. A rate was determined for the vacuum condition, (1.5×10^{-3} $\mu g\ sec.^{-1}\ cm^{-2}$ at 284°C), then helium at a pressure of 0.1 atmosphere was introduced and a new rate determined; (2.17×10^{-4} $\mu g\ sec.^{-1}\ cm^{-2}$) the helium was then pumped out and the vacuum rate determined again and was found unchanged. A slight buoyancy effect, probably caused by convection currents in the helium, caused jumps in the readings on introduction and exhaust of the gas. The jumps were equal and opposite so that the measured evaporation rate with a gas atmosphere was assumed to be correct. The outgassing rate remained constant during the run with 0.1 atmosphere of helium present. The total loss of weight during the test was about 50 μg. The sample was of antimony plated on steel shim stock. The test was carried out at 284°C.

As Fig. 1 shows the evaporation rate of antimony in vacuum at 284°C was approximately 6×10^{-5} $g/cm^2/sec.$ or one twenty-fifth the rate of outgassing in vacuum at this temperature. During the inert atmosphere part of the test at 0.1 atmos. the outgassing rate was reduced by a factor of 7, but, as Fig. 1 shows, the evaporation rate was reduced by a factor of 100 to 1000. Thus, the introduction of the inert gas had the effect of reducing the evaporation rate to a very small fraction of the outgassing rate.

CONCLUSION

An inert gas requires that evaporated molecules diffuse through it and therefore it limits the rate of evaporation. Many evaporated molecules are reflected by the inert gas back to the evaporating surface where they may recondense.

A difference in the accommodation coefficient of this condensing process may explain completely the alteration of the relative rates of outgassing and evaporation of the parent body when an inert atmosphere is introduced. The introduction of such an atmosphere may prove useful when it is desired to outgas other volatile material, for the evaporation rate is probably reduced far more than the outgassing rate in most cases. In the particular case of plated antimony, for the same rate of outgassing, the rate of loss of antimony by evaporation in an inert atmosphere was less than one one-hundredth the rate of loss in a vacuum.[5]

[5] U. S. Patent No. 2,456,968.

A Method of Reducing Deflection Defocusing in Cathode-Ray Tubes

R. G. E. Hutter and S. W. Harrison
Physics Laboratories, Sylvania Electric Products, Inc., Bayside, New York
(Received May 26, 1949)

A discussion of deflection defocusing is given to point out the two-dimensional character of the focusing action of deflection fields. A photograph of spot distortion actually observed is shown. A brief theory of combined spherical and cylindrical lenses is given, and a method of spot correction based on the principle of "beam predistortion" is described. An electron gun used for such correction is shown and experimental results obtained with this gun are exhibited.

INTRODUCTION

BOTH practical experience and theoretical analysis[1–3] show than an electron beam passing through a deflection field is subject to a "two-dimensional" focusing action as well as a bending action. In cathode-ray tubes, such a focusing action is undesirable. If the beam is focused to a small spot at the center of the fluorescent screen and is then deflected, the result of the two-dimensional focusing action will be a distortion of the spot.*

The amount of beam distortion produced by a deflection field varies with the form of the deflection field employed. Attempts to reduce the distortion effects by improving the design of the fields have been relatively successful in the case of magnetic deflection. This has not been true in the case of electrostatic fields where one is more limited in the type of change that can be made in the elements producing the field. The fact that there is a change in the velocity of the electrons as well as a variation in the strength of the electrostatic field also makes defocusing effects in this type of field more inherent than in the magnetic field where there is only a variation in field strength.

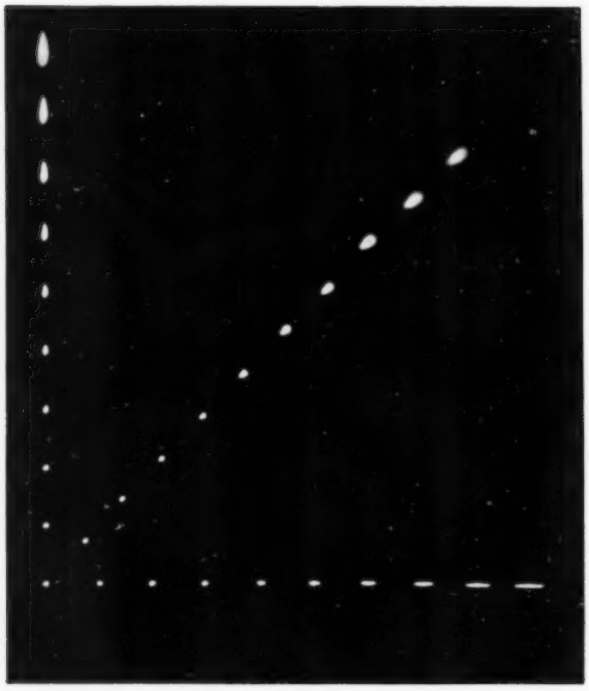


FIG. 1. Spot distortion observed with an electrostatic deflection system of low sensitivity.

The amount of beam distortion associated with any deflection depends also on the parameters of the beam as it enters the field, For a given deflection, a certain set of initial conditions will give a beam which is free of distortion after deflection. Physically, this may be interpreted as saying that the beam is initially distorted in a manner such as to compensate for the distortion produced by the deflection field. To produce such a predistortion, the electron optics of an ordinary axially symmetric focusing system must be modified. The magnitude of this modification will depend on the magnitude and direction of deflection, or, in other words, on the position of the spot on the fluorescent screen.

It is the purpose of this paper to describe a method of spot correction which is based on the principle of beam predistortion. This principle can be employed in both the electrostatic case and the magnetic case, but the particular method described is electrostatic in nature and is used in connection with electrostatic deflection systems.

In describing the method of spot correction it is convenient to begin with a brief discussion of deflection fields. It will be pointed out that the beam distorting action which is associated with deflection is "two-dimensional" in nature. Observed spots will be shown which confirm the character of this type of distortion. This discussion of deflection fields and the effect of such fields on electron beams will be followed by a first-order theory of focusing fields which have spherical and cylindrical lens properties. It will then become apparent that, by combining such focusing fields with deflection fields, the spot distortion effects can be reduced.

Finally, a special electron gun will be described. This

[1] J. Picht and J. Himpan, Ann. d. Physik **39**, 409 (1941); **43**, 53 (1943).
[2] G. Wendt, Die Telefunkenrohre **15**, 100 (1939); Zeits. f. Physik **118**, 593 (1942).
[3] R. G. E. Hutter, J. App. Phys. **18**, 740–758 (1947).
* The properties of a deflection field also cause a non-linear relationship to exist between field strength and deflection. This in turn produces a so-called "pattern distortion." Since this is ordinarily less objectionable, it will be ignored in the present discussion.

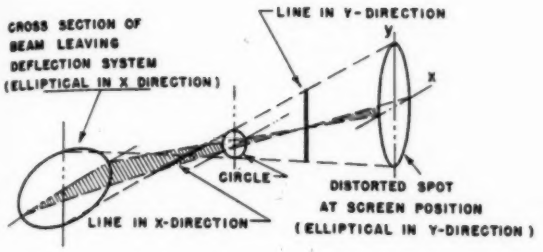

CROSS SECTION OF BEAM LEAVING DEFLECTION SYSTEM (ELLIPTICAL IN X DIRECTION)
LINE IN Y-DIRECTION
y
x
CIRCLE
DISTORTED SPOT AT SCREEN POSITION (ELLIPTICAL IN Y-DIRECTION)
LINE IN X-DIRECTION

FIG. 2. Distortion which results when an electron beam, point focused before deflection, is deflected in the y-direction.

gun is capable of producing an undistorted spot at any point on the screen to which the beam can be deflected. This means that, at the fluorescent screen, the cross section of a deflected beam is circular and has nearly the same diameter as the cross section of the undeflected beam. Photographs taken of the face of a special tube in which this gun was incorporated will show the degree of spot correction achieved.

I. BEAM DISTORTING EFFECTS OF DEFLECTION FIELDS

Actual cathode-ray tubes employing balanced electrostatic deflection fields are designed to have as little spot distortion as possible by keeping the angle of deflection small. In order to study methods of correcting the spot distortion, an electron gun of low sensitivity and hence large distortion effects was designed. A photograph of spots obtained with this gun is shown in Fig. 1 and it can be seen that the general character of the spot distortion is as described by the theory.[1-3]

Both the calculations and the observed results show that the spot is elongated in the direction of deflection. A closer examination of the calculated results, as well as a closer analysis of experimental results for an initially point-focused beam, shows that this characteristic of the spot distortion is due to a focusing effect of the deflection field which is considerably stronger in the direction of deflection than perpendicular to it. The shape of the deflected electron beam on the gun side of the screen is sketched in Fig. 2.

It can be seen from this that in no portion of the beam is there a convergence of all the electron paths

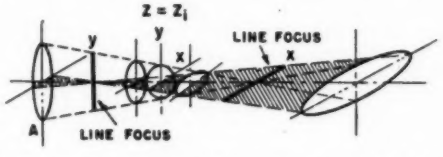


FIG. 3. Shape of an electron beam in the neighborhood of the image plane of a combined spherical and cylindrical electron lens.

to a point. At two places the cross section degenerates into a straight line and at an intermediate point a circular cross section can be found. All other portions of the beam have an elliptical cross section. Reducing the strength of the focusing field of the electron gun will make all paths less convergent, and it is possible to adjust this strength so that the electron beam has a circular cross section at the screen. However, no point focus is now possible in contrast with the point focus assumed at the center of the screen. In the practical case the spot at the center has a finite radius, but no refocusing of the electron lens of rotational symmetry will produce a deflected spot of the same size as the undeflected spot.

A cylindrical lens action must be incorporated in the electron gun to obtain perfect correction for both shape and size. In order to show that this is the case, a brief discussion will be given of the theory of a class of focusing fields which have both spherical and cylindrical lens properties.

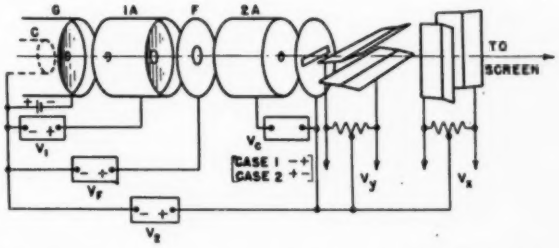


FIG. 4. Electrode arrangement and associated power supplies for an electron gun free of deflection defocusing.

II. THEORY OF COMBINED SPHERICAL AND CYLINDRICAL ELECTRON LENSES

The potential distribution of an electrostatic field which satisfies the symmetry condition,

$$\varphi(x,y,z)=\varphi(-x,y,z)=\varphi(x,-y,z) \tag{1}$$

may be expanded in the series:

$$\varphi(r,z,\theta)=\phi(z)-(1/4)\phi''(z)r^2+\phi_2(z)r^2\cos 2\theta +(1/64)\phi^{IV}(z)r^4-(1/12)\phi_2''(z)r^4\cos 2\theta +\phi_4(z)r^4\cos 4\theta+\cdots. \tag{2}$$

For a field of rotational symmetry $\phi_{2n}(z)=0$. The paraxial ray equations for fields given by Eq. (2) are

$$\left.\begin{aligned} x''+(1/2)(\phi'/\phi)x'+(1/4)(\phi''/\phi)x&=(\phi_2/\phi)x \\ y''+(1/2)(\phi'/\phi)y'+(1/4)(\phi''/\phi)y&=-(\phi_2/\phi)y \end{aligned}\right\}. \tag{3}$$

These equations may be integrated under the following conditions:

(1) The general solution of the paraxial ray equation is known for the case $\phi_{2n}(z)=0$ (spherical lens).

(2) Terms which contain $\phi_{2n}(z)$ (combined spherical and cylindrical lens) may be considered small distortion terms.

With these conditions, the integration of Eq. (3) may be carried out by employing the methods ordinarily used in lens aberration theory.

Let $r(z)$ be the general solution of the paraxial ray equation:

$$r''+(1/2)(\phi'/\phi)r'+(1/4)(\phi''/\phi)r=0. \tag{4}$$

If $r(z)$ has the value r_0 in the object plane $z=z_0$ and

the value r_a in the aperture plane $z=z_a$, the solution may be written

$$r(z)=r_0 r_\gamma(z)+(r_a/r_{aa})r_a(z), \tag{5}$$

where $r_a(z)$ and $r_\gamma(z)$ are two particular solutions which

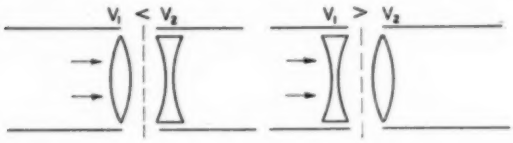


FIG. 5. A simple two-cylinder rotationally symmetric electron lens and the optical analogies for $V_1<V_2$ and $V_1>V_2$.

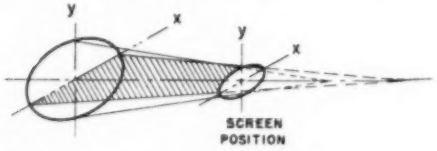


FIG. 6. Distortion of an undeflected electron beam necessary to compensate for the focusing action associated with a deflection in the x-direction.

satisfy the conditions

$$\begin{aligned} r_a(z_0)&=0 \quad r_a'(z_0)=1 \\ r_\gamma(z_0)&=1 \quad r_\gamma(z_a)=0. \end{aligned} \tag{6}$$

The values of $x(z)$ and $y(z)$ of Eq. (3) in the image plane are then given** by:

$$\left.\begin{aligned} x(z_i)=\Delta x_i&=-\frac{M}{(\phi_0)^{\frac{1}{4}}}\int_{z_0}^{z_i}\frac{\phi_2(z)}{(\phi(z))^{\frac{1}{4}}}r(z)r_a(z)dz \\ y(z_i)=\Delta y_i&=+\frac{M}{(\phi_0)^{\frac{1}{4}}}\int_{z_0}^{z_i}\frac{\phi_2(z)}{(\phi(z))^{\frac{1}{4}}}r(z)r_a(z)dz \end{aligned}\right\}. \tag{7}$$

For electron paths for which $r_0=0$, Eqs. (5) and (7) may be combined to give

$$\left.\begin{aligned} \Delta x_i&=-\frac{M}{(\phi_0)^{\frac{1}{4}}}\frac{r_a}{r_{aa}}\int_{z_0}^{z_i}\frac{\phi_2(z)}{(\phi(z))^{\frac{1}{4}}}r_a^2(z)dz \\ \Delta y_i&=+\frac{M}{(\phi_0)^{\frac{1}{4}}}\frac{r_a}{r_{aa}}\int_{z_0}^{z_i}\frac{\phi_2(z)}{(\phi(z))^{\frac{1}{4}}}r_a^2(z)dz \end{aligned}\right\}, \tag{8}$$

where $\phi_0=\phi(z_0)$ and $M=r_\gamma(z_i)$ is the magnification of the rotationally symmetric lens. Equation (8) shows that $\Delta x_i=-\Delta y_i$ and the cross section of the electron beam in the neighborhood of the image plane will vary as is shown in Fig. 3.

The type of lens described above may be used to produce a beam which converges differently in the x- and y-direction. This is the effect which was observed in the defocusing action of a deflection field. Such a lens may thus be used to compensate for the asymmetrical focusing action of a deflection system.

** See Madelung, *Die Mathematischen Hilfsmittel des Physikers* (Dover Publications, New York, 1943), third edition, p. 183.

III. AN ELECTRON GUN FREE OF DEFLECTION DEFOCUSING

On the basis of the analysis given above, it is possible to design an electron gun which is free of deflection defocusing. Such a gun is shown schematically in Fig. 4.

With the exception of a slit electrode, this gun is similar to the electron gun used in Type 7GP4 cathode-ray tubes. The slit is parallel to the first pair of deflection plates. The first and second anode (1A and 2A) are usually operated at the same high potential with respect to the cathode (i.e., $V_1=V_2$, $V_c=0$). The focusing electrode (F) is operated at a fraction of the value of the anode potential and the voltage of F is adjusted to produce as small a spot as possible on the screen. Because of space charge effects, velocity distribution effects, and lens aberrations, this spot is not a true point focus. The deflection systems are balanced as indicated in Fig. 4. The space between the last aperture in the second anode and the screen is at the uniform potential of the last anode if the deflection potentials V_x and V_y are equal to zero.

The addition of the slit electrode will not change anything when $V_c=0$ and $V_1=V_2$ provided the slit does not intercept electrons and does not distort the field in the neighborhood of the last aperture. If, however, $V_c\neq 0$, the field between an electrode of circular symmetry (2A) and one of two-dimensional symmetry (the slit electrode) will constitute a lens similar to that discussed in Section II of this paper.

In order to understand the performance of this modified gun in connection with spot correction, it should first be noted that the sign of the correction voltage V_c may be reversed. It can be reasoned that this reverses the sign of $\phi_2(z)$ in Eqs. (3) and (8). Now when $\phi_2(z)$ is positive, Δx_i is negative and Δy_i positive,

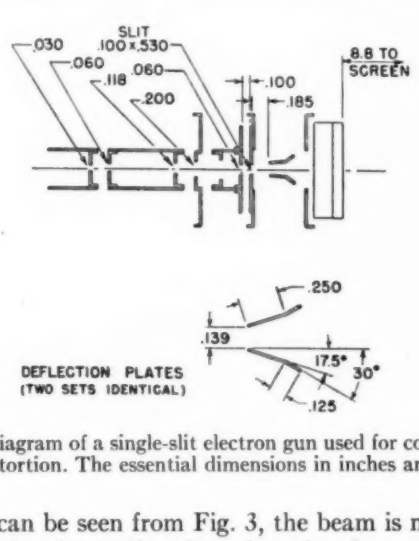


FIG. 7. Diagram of a single-slit electron gun used for correction of spot distortion. The essential dimensions in inches are shown.

and, as can be seen from Fig. 3, the beam is more convergent in the x-direction than in the y-direction. Similarly, with $\phi_2(z)$ negative, the convergence is greater in the y-direction. This reasoning presupposes a knowledge of the field distribution; since the field is

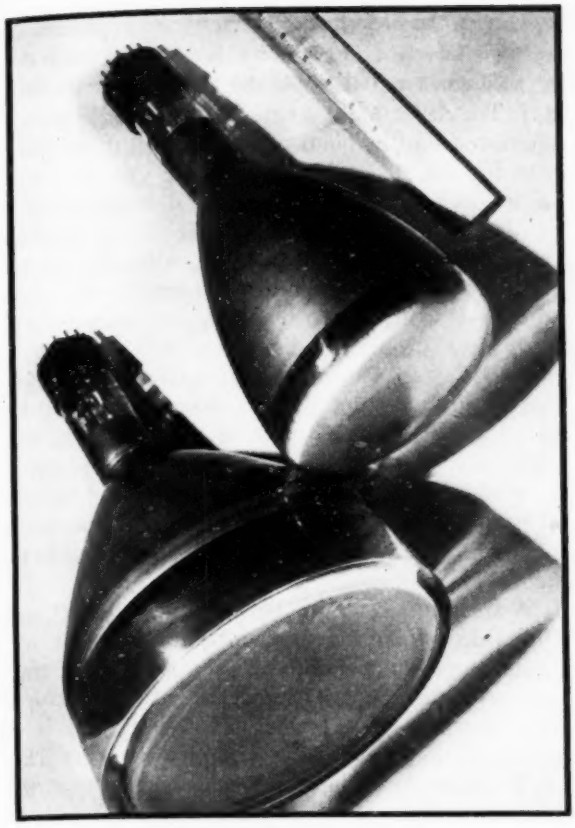


FIG. 8. Special 10-inch tube with single-slit gun and standard 7GP4 electrostatic tube showing over-all length the same for both tubes.

rather complex it is easier to proceed with a light-optical analogy.

Such analogies are often used in electron optics. Take, for instance, a simple rotationally symmetric lens which exists between two cylinders at different potentials. Such a lens is shown in Fig. 5. Although the net effect of this lens is a converging action, it should be pointed out that part of the lens has a diverging action and part a converging action. The regions in which these actions exist will depend on the potentials on the two cylinders and the two possible cases are shown, with the light optical analogies, in Fig. 5. The same thing may be said of apertured disks which separate two regions of uniform potential; this analogy holds if one of the rotationally symmetric electrodes is replaced by a slit electrode. In the latter case, of course, the lens action in the region of the slit electrode will be two-dimensional.

The performance of the modified gun may now be explained. Consider first a deflection in the y-direction and the correction for the spot distortion associated with such a deflection. With $V_c=0$, V_1, V_2 and V_f are adjusted so that the beam is "point-focused" at the center of the screen. (V_1 may or may not be equal to V_2). If the polarity of V_c is then as shown for case I in Fig. 4, an increase of V_c will mean:

(1) A reduction in the strength of the rotationally symmetric lens associated with $1A$, F, and $2A$;

(2) The creation of a "two-dimensional" diverging lens near the slit electrode and a converging, rotationally symmetric lens near the last aperture of $2A$.

With the possibility of varying V_f which changes the strength of the spherical lens, and V_c, which changes the relative convergence in the x- and y-directions, it is possible to distort the beam in such a way that its cross section at the screen position is similar to that shown at A in Fig. 3. It follows that the beam converges less in the y-direction than in the x-direction and that the rays in both the yz- and xz-planes intersect the screen before they cross the z axis. Since a given predistortion of this nature is necessary to compensate for the focusing action associated with a deflection in the y-direction, it is possible to adjust the spot at the center of the screen in such a way that the beam will be point-focused at the screen when deflected. In practice, the procedure is to deflect the beam, observe the spot distortion, and then adjust V_f and V_c so that the spot distortion is eliminated.

Since the electron gun also provides a deflection in the x-direction, it is necessary to provide for spot correction in this direction too. For this, another slit electrode following the first slit electrode with the slit rotated through 90° can be used with nearly the same potential arrangement. However, it is possible to use the single slit electrode shown in Fig. 4 for correction in both the x- and y-directions as will be shown. The polarity of V_c is chosen as illustrated for case II in Fig. 4. Now an increase of V_c will mean:

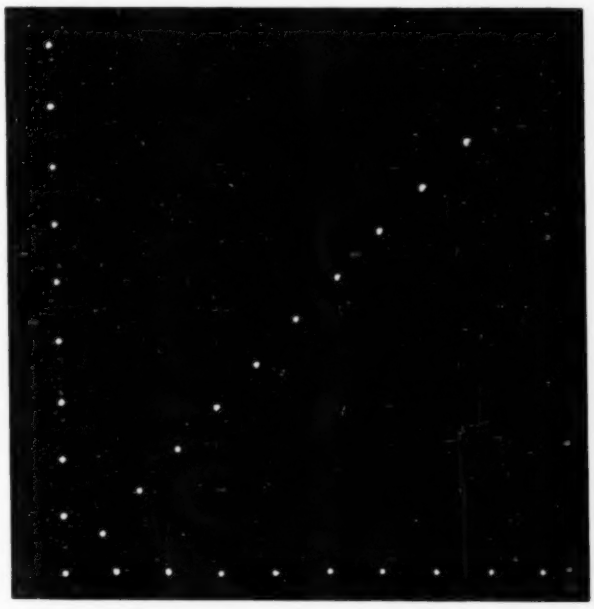


FIG. 9. Corrected spots on special 10-inch electrostatic tube with proper predistortion of the beam. The spots before correction were shown in Fig. 1.

(1) An increase in the strength of the rotationally symmetric lens associated with $1A$, F, and $2A$;

(2) The creation of a "two-dimensional" converging lens near the slit electrode and a diverging rotationally symmetric lens near the last aperture of $2A$.

Again, the possibility of varying V_f and V_c means that the beam can be predistorted, but in this case the two-dimensional lens is converging in the y-direction and the resulting beam will be less convergent in the x-direction than in the y-direction as illustrated in Fig. 6. This is the type of predistortion necessary to compensate for the focusing action associated with a deflection in the x-direction, and again the spot distortion of a deflected beam can be eliminated by adjusting V_f and V_c.

Since the spot in a cathode-ray tube is brought to an arbitrary position on the screen by deflection in both the x- and y-directions, there is some focusing in each direction due to the deflection fields. Equal amounts of focusing in each direction may be compensated for by weakening the axially symmetric focus, V_f. If the focusing action is stronger in one direction than in the other, the excess is treated just as a focusing action in one direction. Physically, this means that, for any position of the spot on the screen, it is always possible to find a combination of V_c and ΔV_f which produces a corrected spot.

IV. EXPERIMENTAL RESULTS

In an effort to check the analysis experimentally, a gun was constructed with the dimensions shown in Fig. 7. The deflection system of this gun was purposely made poor from the point of view of spot distortion by shortening the deflection plates. The gun was sealed in a ten-inch bulb with an over-all length equal to that of a standard Type 7GP4 tube. The special ten-inch tube and the 7GP4 are shown in Fig. 8. The spot distortion which results when the special tube is operated with $V_c=0$ was shown in Fig. 1. The results obtained after proper predistortion of the beam are shown in Fig. 9. The functional relationship between V_c, ΔV_f, and the deflection voltages are plotted in Figs. 10 and 11. Voltages must be applied to both pairs of deflection plates to produce a deflection along the 45° radius and, in this case, the "deflection voltage" read from the abscissa in Fig. 11 is the square root of the sum of the squares of the x- and y-deflection voltages.

It is apparent from Fig. 9 that nearly perfect spots are obtained when the predistortion is applied, even for deflections which normally produce a considerable amount of distortion. Commercial tubes have, of course, much less deflection defocusing than is obtained with the special tube without correction. However, if the angle of deflection in a regular tube were

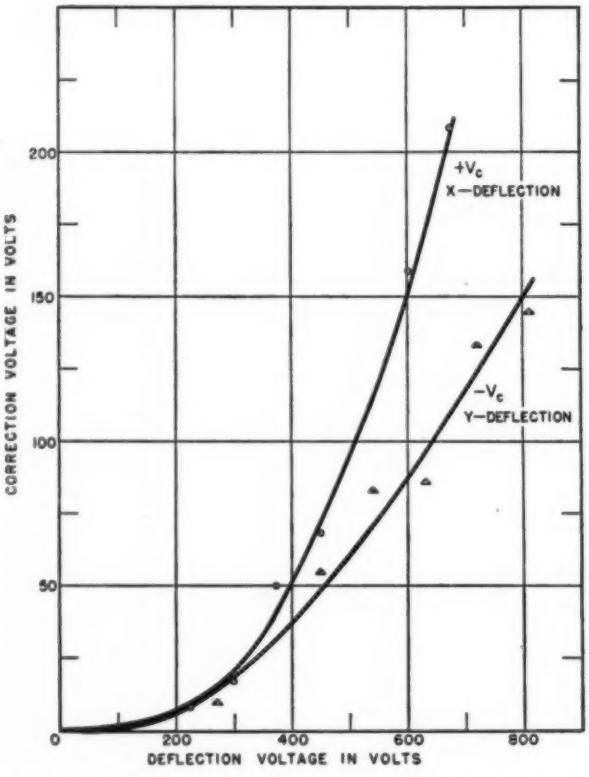


FIG. 10. Variation of correction voltages with deflection voltage for correction of spot distortion as shown in Fig. 9. The last anode voltage was 2000 volts and the grid bias (−162) volts with a pulse of 20 μsec., 110 cycles.

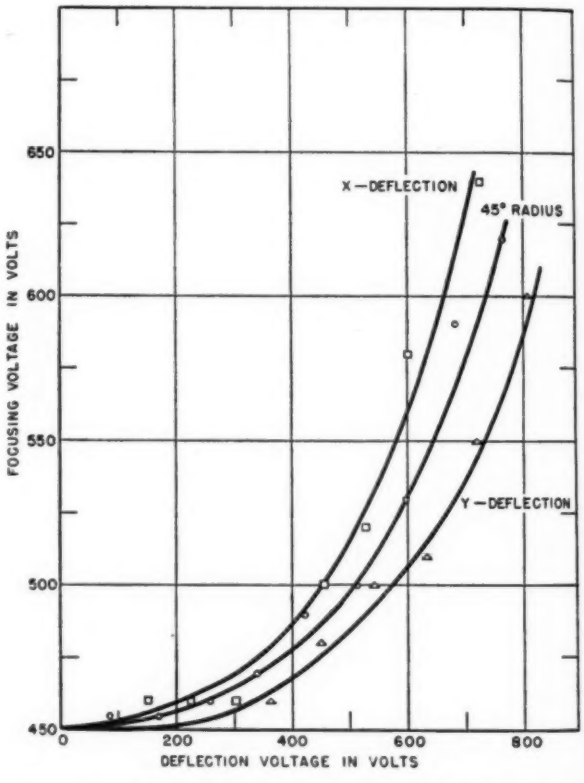


FIG. 11. Variation of focusing voltage with deflection voltage for correction of spot distortion as shown in Fig. 9. The last anode voltage was 2000 volts and the grid bias (−162) volts with a pulse of 20 μsec., 110 cycles.

increased in order to reduce the length of the tube, the spot distortion would be quite objectionable. In such a situation correction by predistortion would be very effective.

C. CONCLUDING REMARKS

The discussion above has been limited to the case of a tube employing a static deflection. In the operation of a cathode-ray tube under normal circumstances, circuits which permit automatic correction during the scanning cycle are required but will not be discussed here. The magnitudes of the correction potentials and their functional relationships to the deflection potentials depend on the particular design of the focusing and deflection systems and upon the degree of spot correction desired.

Although the correction method discussed here is for an electrostatic focusing and deflection system, it is possible to apply the principle of beam predistortion to magnetic focusing and deflection systems as well. In this connection, however, it should be pointed out that magnetic lenses, either spherical or cylindrical, are everywhere converging, i.e., there is no diverging region. If spot correction is attempted by the use of "two-dimensional" magnetic lenses, two such lenses rotated 90° with respect to each other are required.

ACKNOWLEDGMENT

The authors wish to take this opportunity of acknowledging the valuable contributions made by L. H. McKee of these laboratories in carrying out most of the experimental part of the work described.

Transmission of Elastic Waves through a Stratified Solid Medium

WILLIAM T. THOMSON*
University of Wisconsin, Madison, Wisconsin
(Received June 14, 1949)

The transmission of a plane elastic wave at oblique incidence through a stratified solid medium consisting of any number of parallel plates of different material and thickness is studied theoretically. The matrix method is used to systematize the analysis and to present the equations in a form suitable for computation.

INTRODUCTION

THE transmission of elastic waves in a homogeneous solid has been extensively treated by G. G. Stokes.[1] Barring surface waves of the type discussed by Rayleigh,[2] such waves are propagated as dilatation and rotation waves (frequently referred to as shear waves), each traveling with a different speed. It is thus necessary in dealing with such problems first to express the particle displacements in terms of the dilatation and rotation.[3]

The transmission of plane elastic waves through a stratified medium consisting of alternate parallel planes of solid and liquid layers has been treated by R. B. Lindsay[4] and others.[5,6] Since ideal fluids are incapable of sustaining tangential forces, the shearing stresses in the solid-liquid strata must vanish at the surfaces of the solid plates, thus reducing the problem to that of a single plate.

When the stratified medium consists of parallel solid plates without liquid layers between them, the problem cannot be reduced to that of a single section. The equations for one section must now be related to those of the adjoining section by the continuity of particle velocity, normal and shearing stresses at the interface. This is most conveniently accomplished by matrices which avoids unwieldly mathematical expressions and facilitates the computational task. The direction of propagation through each strata is established by Snell's law.

Nomenclature

ξ, η, ζ = particle displacement in x, y, z directions
$\dot{\xi}, \dot{\eta}, \dot{\zeta}$ = particle velocity in x, y, z directions
Δ = dilatation
ω = rotation (subscript indicates direction of normal to rotation plane)
$c_\Delta = [(\lambda+2G)/\rho]^{\frac{1}{2}}$ = velocity of dilatational wave in solid
$c_\omega = (G/\rho)^{\frac{1}{2}}$ = velocity of rotation or shear wave in solid
c_0 = velocity of propagation in the fluid
$\lambda = \nu E/[(1+\nu)(1-2\nu)]$
E = Young's modulus of elasticity
ν = Poisson's ratio
G = modulus of elasticity in shear

* Associate Professor of Mechanics.
[1] G. G. Stokes, "Dynamic Theory of Diffraction," *Mathematical and Physical Papers* (Cambridge University Press, London, 1883), Vol. II, part 1, pp. 250–290.
[2] Love, *Mathematical Theory of Elasticity* (Dover Publications, New York, 1944), pp. 307–309.
[3] Lord Rayleigh, *Theory of Sound* (Dover Publications, New York, 1945), part II, p. 417.
[4] R. B. Lindsay, J. Acous. Soc. Am. **11**, 178–183 (1939).
[5] J. B. Smyth and R. B. Lindsay, J. Acous. Soc. Am. **16**, 20–25 (1944).
[6] H. Reissner, "Der senkrechte und schrage Durchtritt einer in einem flüssigen Medium erzeugten ebenen Dilatations Welle durch einen in diesem Medium befindliche planparallele feste Platte," Helv. Phys. Acta. **11**, 140–155 (1938).

ρ = mass density
p = frequency in radians/sec.
$h = p/c_\Delta$
$k = p/c_\omega$
σ = normal stress
τ = shearing stress (subscript indicates direction of normal to plane of shear)
l, m, n = direction cosines of normal to advancing wave with respect to x, y, z axes
d = thickness of plates
$P = hn_\Delta d$
$Q = kn_\omega d$
R = reflection coefficient
T = transmission coefficient
$\varphi', \varphi'', \varphi'''$ = incident, reflected and transmitted waves in fluid surrounding the stratified medium
θ_0 = incident angle.

GENERAL EQUATIONS FOR DISPLACEMENTS IN TERMS OF DILATATION AND ROTATION

The dilatation and rotation are defined by the following equations.

$$\Delta = \frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \tag{1}$$

$$\left.\begin{aligned} \omega_x &= \tfrac{1}{2}\left(\frac{\partial \zeta}{\partial y} - \frac{\partial \eta}{\partial z}\right) \\ \omega_y &= \tfrac{1}{2}\left(\frac{\partial \xi}{\partial z} - \frac{\partial \zeta}{\partial x}\right) \\ \omega_z &= \tfrac{1}{2}\left(\frac{\partial \eta}{\partial x} - \frac{\partial \xi}{\partial y}\right) \end{aligned}\right\}. \tag{2}$$

Differentiating Δ with respect to x, y, and z, we obtain the three equations

$$\left.\begin{aligned} \nabla^2\xi &= \frac{\partial \Delta}{\partial x} - 2\frac{\partial \omega_z}{\partial y} + 2\frac{\partial \omega_y}{\partial z}, \\ \nabla^2\eta &= \frac{\partial \Delta}{\partial y} - 2\frac{\partial \omega_z}{\partial z} + 2\frac{\partial \omega_z}{\partial x}, \\ \nabla^2\zeta &= \frac{\partial \Delta}{\partial z} - 2\frac{\partial \omega_y}{\partial x} + 2\frac{\partial \omega_x}{\partial y}. \end{aligned}\right\} \tag{3}$$

The general solution for the displacements in terms of Δ and ω which also satisfies the wave equation for dilation and rotation of harmonic time variation is

$$\left.\begin{aligned} \xi &= -\frac{1}{h^2}\frac{\partial \Delta}{\partial x} + \frac{2}{k^2}\frac{\partial \omega_z}{\partial y} - \frac{2}{k^2}\frac{\partial \omega_y}{\partial z}, \\ \eta &= -\frac{1}{h^2}\frac{\partial \Delta}{\partial y} + \frac{2}{k^2}\frac{\partial \omega_z}{\partial z} - \frac{2}{k^2}\frac{\partial \omega_z}{\partial x}, \\ \zeta &= -\frac{1}{h^2}\frac{\partial \Delta}{\partial z} + \frac{2}{k^2}\frac{\partial \omega_y}{\partial x} - \frac{2}{k^2}\frac{\partial \omega_x}{\partial y}. \end{aligned}\right\} \tag{4}$$

For instance, from the first of Eq. (4)

$$\nabla^2\xi = -\frac{1}{h^2}\frac{\partial}{\partial x}(\nabla^2\Delta) + \frac{2}{k^2}\frac{\partial}{\partial y}(\nabla^2\omega_z) - \frac{2}{k^2}\frac{\partial}{\partial z}(\nabla^2\omega_y)$$

which compared with the first of Eq. (3) results in the wave equations

$$\begin{aligned} (\nabla^2 + h^2)\Delta &= 0, \\ (\nabla^2 + k^2)\omega_x &= 0, \\ (\nabla^2 + k^2)\omega_y &= 0. \end{aligned}$$

GENERAL EQUATIONS FOR STRESSES IN TERMS OF DILATATION AND ROTATION

From the theory of elasticity we have the equations for stress in terms of displacements as follows:

$$\sigma_x = \lambda\Delta + 2G(\partial\xi/\partial x)$$

and two others,

$$\tau_x = G\left(\frac{\partial \eta}{\partial z} + \frac{\partial \zeta}{\partial y}\right)$$

and two others. Hence by substituting from Eq. (4) we obtain the equations for stress in terms of ∇ and ω

$$\sigma_x = \lambda\Delta + 2G\left(-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial x^2} + \frac{2}{k^2}\frac{\partial^2\omega_z}{\partial x\partial y} - \frac{2}{k^2}\frac{\partial^2\omega_y}{\partial x\partial z}\right)$$

$$\sigma_y = \lambda\Delta + 2G\left(-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial y^2} + \frac{2}{k^2}\frac{\partial^2\omega_x}{\partial y\partial z} - \frac{2}{k^2}\frac{\partial^2\omega_z}{\partial x\partial y}\right)$$

$$\sigma_z = \lambda\Delta + 2G\left(-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial z^2} + \frac{2}{k^2}\frac{\partial^2\omega_y}{\partial x\partial z} - \frac{2}{k^2}\frac{\partial^2\omega_x}{\partial y\partial z}\right)$$

$$\begin{aligned} \tau_x = 2G\Bigg(&-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial y\partial z} - \frac{1}{k^2}\frac{\partial^2\omega_x}{\partial x\partial z} - \frac{1}{k^2}\frac{\partial^2\omega_x}{\partial y^2} \\ &+ \frac{1}{k^2}\frac{\partial^2\omega_z}{\partial z^2} + \frac{1}{k^2}\frac{\partial^2\omega_y}{\partial x\partial y}\Bigg) \end{aligned} \tag{5}$$

$$\begin{aligned} \tau_y = 2G\Bigg(&-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial x\partial z} - \frac{1}{k^2}\frac{\partial^2\omega_z}{\partial x\partial y} - \frac{1}{k^2}\frac{\partial^2\omega_y}{\partial z^2} \\ &+ \frac{1}{k^2}\frac{\partial^2\omega_y}{\partial x^2} + \frac{1}{k^2}\frac{\partial^2\omega_z}{\partial y\partial z}\Bigg) \end{aligned}$$

$$\begin{aligned} \tau_z = 2G\Bigg(&-\frac{1}{h^2}\frac{\partial^2\Delta}{\partial x\partial y} - \frac{1}{k^2}\frac{\partial^2\omega_y}{\partial y\partial z} - \frac{1}{k^2}\frac{\partial^2\omega_z}{\partial x^2} \\ &+ \frac{1}{k^2}\frac{\partial^2\omega_z}{\partial y^2} + \frac{1}{k^2}\frac{\partial^2\omega_x}{\partial x\partial z}\Bigg). \end{aligned}$$

TRANSMISSION OF PLANE WAVES THROUGH STRATIFIED SOLID PLATES

Choosing the coordinate axes as shown in Fig. 1, ω_x, ω_y and m will be zero. Considering the nth plate, the dilitation and rotation satisfying the wave equation are

$$\begin{aligned}\Delta&=[\Delta' \exp(ih(l_\Delta x+n_\Delta z))+\Delta'' \exp(ih(l_\Delta x-n_\Delta z))]e^{ipt}\\ \omega&=[\omega' \exp(ik(l_\omega x+n_\omega z))+\omega'' \exp(ik(l_\omega x-n_\omega z))]e^{ipt}\end{aligned} \tag{6}$$

where the primed and double primed quantities represent the advancing and reflected waves. Substituting into Eq. (4) and (5) and omitting the trivial time factor, we obtain the following equations at $x=0$, $z=d$, and $x=0$, $z=0$, expressed in matrix form.

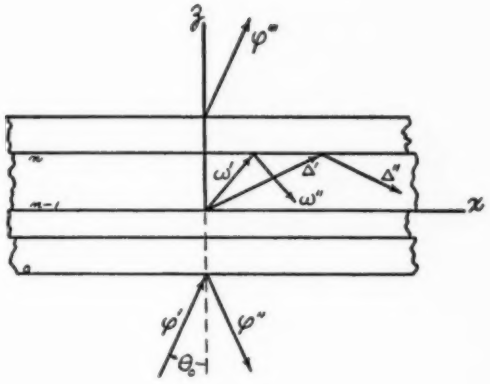


FIG. 1.

$$\begin{pmatrix}\dot{\xi}_n\\ \dot{\zeta}_n\\ \sigma_n\\ \dfrac{\tau_n}{2G}\end{pmatrix}=\begin{pmatrix}\dfrac{pl_\Delta}{h}\cos P & \dfrac{ipl_\Delta}{h}\sin P & \dfrac{2pn_\omega}{k}\cos Q & \dfrac{i2pn_\omega}{k}\sin Q\\ \dfrac{ipn_\Delta}{h}\sin P & \dfrac{pn_\Delta}{h}\cos P & -\dfrac{i2pl_\omega}{k}\sin Q & -\dfrac{2pl_\omega}{k}\cos Q\\ (\lambda+2Gn_\Delta^2)\cos P & i(\lambda+2Gn_\Delta^2)\sin P & -4Gl_\omega n_\omega\cos Q & -i4Gl_\omega n_\omega\sin Q\\ il_\Delta n_\Delta\sin P & l_\Delta n_\Delta\cos P & -i(l_\omega^2-n_\omega^2)\sin Q & -(l_\omega^2-n_\omega^2)\cos Q\end{pmatrix}\begin{pmatrix}\Delta'+\Delta''\\ \Delta'-\Delta''\\ \omega'-\omega''\\ \omega'+\omega''\end{pmatrix}, \tag{7}$$

$$\begin{pmatrix}\dot{\xi}_{n-1}\\ \dot{\zeta}_{n-1}\\ \sigma_{n-1}\\ \dfrac{\tau_{n-1}}{2G}\end{pmatrix}=\begin{pmatrix}\dfrac{pl_\Delta}{h} & 0 & \dfrac{2pn_\omega}{k} & 0\\ 0 & \dfrac{pn_\Delta}{h} & 0 & -\dfrac{2pl_\omega}{k}\\ (\lambda+2Gn_\Delta^2) & 0 & -4Gl_\omega n_\omega & 0\\ 0 & l_\Delta n_\Delta & 0 & -(l_\omega^2-n_\omega^2)\end{pmatrix}\begin{pmatrix}\Delta'+\Delta''\\ \Delta'-\Delta''\\ \omega'-\omega''\\ \omega'+\omega''\end{pmatrix}. \tag{8}$$

The inversion of Eq. (8) is

$$\begin{pmatrix}\Delta'+\Delta''\\ \Delta'-\Delta''\\ \omega'-\omega''\\ \omega'+\omega''\end{pmatrix}=\begin{pmatrix}-\dfrac{4Gl_\omega n_\omega}{D_1} & 0 & -\dfrac{2pn_\omega}{kD_1} & 0\\ 0 & -\dfrac{(l_\omega^2-n_\omega^2)}{D_2} & 0 & \dfrac{2pl_\omega}{kD_2}\\ -\dfrac{(\lambda+2Gn_\Delta^2)}{D_1} & 0 & \dfrac{pl_\Delta}{hD_1} & 0\\ 0 & -\dfrac{l_\Delta n_\Delta}{D_2} & 0 & \dfrac{pn_\Delta}{hD_2}\end{pmatrix}\begin{pmatrix}\dot{\xi}_{n-1}\\ \dot{\zeta}_{n-1}\\ \sigma_{n-1}\\ \dfrac{\tau_{n-1}}{2G}\end{pmatrix} \tag{9}$$

where

$$D_1=-\left[\frac{2pn_\omega}{k}(\lambda+2Gn_\Delta^2)+\frac{pl_\Delta}{h}4Gl_\omega n_\omega\right],$$
$$D_2=-\left[\frac{pn_\Delta}{h}(l_\omega^2-n_\omega^2)-\frac{2pl_\omega}{k}l_\Delta n_\Delta\right]. \tag{10}$$

Thus by substituting Eq. (9) in (7) we obtain the recurrence formula

$$\begin{pmatrix}\xi_n\\ \zeta_n\\ \sigma_n\\ \dfrac{\tau_n}{2G}\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&a_{13}&a_{14}\\ a_{21}&a_{22}&a_{23}&a_{24}\\ a_{31}&a_{32}&a_{33}&a_{34}\\ a_{41}&a_{42}&a_{43}&a_{44}\end{pmatrix}\begin{pmatrix}\xi_{n-1}\\ \zeta_{n-1}\\ \sigma_{n-1}\\ \dfrac{\tau_{n-1}}{2G}\end{pmatrix} \tag{11}$$

where the expressions for the matrix elements are given in the Appendix.

The continuity of particle velocities and stresses at the interface enables one to express the quantities at the nth plate in terms of corresponding quantities at the first plate. Starting with Eq. (11) we can write

$$\begin{pmatrix}\xi_n\\ \zeta_n\\ \sigma_n\\ \dfrac{\tau_n}{2G}\end{pmatrix}=\begin{pmatrix}\\ a_n\\ \\ \end{pmatrix}\begin{pmatrix}\\ a_{n-1}\\ \\ \end{pmatrix}\cdots\begin{pmatrix}\\ a_1\\ \\ \end{pmatrix}\begin{pmatrix}\xi_0\\ \zeta_0\\ \sigma_0\\ \dfrac{\tau_0}{2G}\end{pmatrix}$$
$$=\begin{pmatrix}A_{11}&A_{12}&A_{13}&A_{14}\\ A_{21}&A_{22}&A_{23}&A_{24}\\ A_{31}&A_{32}&A_{33}&A_{34}\\ A_{41}&A_{42}&A_{43}&A_{44}\end{pmatrix}\begin{pmatrix}\xi_0\\ \zeta_0\\ \sigma_0\\ \dfrac{\tau_0}{2G}\end{pmatrix} \tag{12}$$

where $[a_k]$ represents the square matrix of Eq. (11) for the kth plate.

The direction cosines l_Δ, n_Δ, l_ω and n_ω for each plate can be obtained from Snell's law.

$$\frac{\sin\theta_0}{c_0}=\frac{\sin\theta_\Delta}{c_\Delta}=\frac{\sin\theta_\omega}{c_\omega}. \tag{13}$$

Letting θ_1 and θ_2 be the critical angles for the dilation and shear waves respectively, it is apparent that for any plate there will be three regions for θ_0.

(1) When $0\leq\theta_0\leq\theta_1$, both θ_Δ and θ_ω will be real.
(2) When $\theta_1<\theta_0\leq\theta_2$, θ_Δ will be complex and θ_ω real.
(3) When $\theta_2<\theta_0\leq\pi/2$, both θ_Δ and θ_ω will be complex.

As an example, in region (2) we let $\theta_\Delta=\alpha+i\beta$ and obtain $\alpha=\pi/2$ and $\cosh\beta=(c_\Delta/c_0)\sin\theta_0$. Thus we replace $\sin\theta_\Delta$ and $\cos\theta_\Delta$ by $\cosh\beta$ and $i\sinh\beta$ respectively.

If the planes 0 and n terminate in a liquid medium, the shearing stresses τ_0 and τ_n must vanish, resulting in the boundary equation,

$$A_{41}\xi_0+A_{42}\zeta_0+A_{43}\sigma_0=0. \tag{14}$$

Substituting Eq. (14) into Eq. (12) with $\tau_0=\tau_n=0$

$$\begin{pmatrix}\zeta_n\\ \sigma_n\end{pmatrix}=\begin{vmatrix}\left(A_{22}-\dfrac{A_{21}A_{42}}{A_{41}}\right)&\left(A_{23}-\dfrac{A_{21}A_{43}}{A_{41}}\right)\\ \left(A_{32}-\dfrac{A_{31}A_{42}}{A_{41}}\right)&\left(A_{33}-\dfrac{A_{31}A_{43}}{A_{41}}\right)\end{vmatrix}\begin{pmatrix}\zeta_0\\ \sigma_0\end{pmatrix}. \tag{15}$$

To determine the transmission and reflection, we have from Eq. (6) with $\omega=0$ for the liquid medium, the following.

$$\zeta_n=c_0n_0\varphi'''\qquad \zeta_0=c_0n_0(\varphi'-\varphi'')$$
$$\sigma_n=-c_0^2\rho_0\varphi'''\qquad \sigma_0=-c_0^2\rho_0(\varphi'+\varphi'') \tag{16}$$

where φ', φ'' and φ''' represent the incident, reflected, and transmitted waves respectively. Substituting into Eq. (15) the reflection and transmission coefficients become,

$$R=\left|\frac{\varphi''}{\varphi'}\right|^2=\left|\frac{[A_{22}-(A_{21}A_{42}/A_{41})]-[A_{33}-(A_{31}A_{43}/A_{41})]-[A_{23}-(A_{21}A_{43}/A_{41})](c_0\rho_0/n_0)+[A_{32}-(A_{31}A_{42}/A_{41})](n_0/c_0\rho_0)}{[A_{22}-(A_{21}A_{42}/A_{41})]+[A_{33}-(A_{31}A_{43}/A_{41})]+[A_{23}-(A_{21}A_{43}/A_{41})](c_0\rho_0/n_0)+[A_{32}-(A_{31}A_{42}/A_{41})](n_0/c_0\rho_0)}\right|^2, \tag{17}$$

$$T=1-R. \tag{18}$$

SPECIAL CASES

The following special cases are of interest.

(1) Stratified medium is fluid, thus all shearing stresses are zero. Letting $G=Q=c_\omega=0$, we find that the 1st and 4th columns of $[a_k]$ are zero. This requires that all terms in $[A]$ vanish except the four central terms and we obtain the equations

$$\begin{bmatrix}\zeta_n\\ \sigma_n\end{bmatrix}=\begin{bmatrix}A_{22}&A_{23}\\ A_{32}&A_{33}\end{bmatrix}\begin{bmatrix}\zeta_0\\ \sigma_0\end{bmatrix} \tag{19}$$

$$R=\left|\frac{A_{22}-A_{33}-A_{23}(c_0\rho_0/n_0)+A_{32}(n_0/c_0\rho_0)}{A_{22}+A_{33}+A_{23}(c_0\rho_0/n_0)+A_{32}(n_0/c_0\rho_0)}\right|^2, \quad (20)$$

$$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} \cosh nd & \frac{in}{c\rho}\sinh nd \\ \frac{ic\rho}{n}\sinh nd & \cosh nd \end{vmatrix}. \quad (21)$$

Thus for a single liquid layer we obtain the well-known relation (see Rayleigh, Part II, Eq. (10), p. 88)

$$R=\frac{[(n_1/n_0)(c_0\rho_0/c_1\rho_1)-(n_0/n_1)(c_1\rho_1/c_0\rho_0)]^2}{4\cot^2 h_1n_1d_1+[(n_1/n_0)(c_0\rho_0/c_1\rho_1)+(n_0/n_1)(c_1\rho_1/c_0\rho_0)]^2}. \quad (22)$$

(2) Stratified medium is solid, and the incident ray is normal to the plates. In this case the shears vanish and Eqs. (20), (21), and (22) become applicable with $n_0=n_1=1$.

(3) Transmission through a single solid plate at the critical angles.

(*a*) At the critical angle θ_1 for the dilation wave $l_\Delta=1$ and $n_\Delta=0$. Re-examination of Eqs. (7) and (8) shows that every term in the second row is zero which requires that the terms in the second row of Eq. (11) be also zero. Furthermore, Eq. (15) becomes

$$\begin{vmatrix} \xi_1 \\ \sigma_1 \end{vmatrix} = \begin{vmatrix} 0 & \left(a_{23}-a_{21}\frac{a_{43}}{a_{41}}\right) \\ 0 & \left(a_{33}-a_{31}\frac{a_{43}}{a_{41}}\right) \end{vmatrix} \begin{vmatrix} \xi_0 \\ \sigma_0 \end{vmatrix} \quad (23)$$

and the ratio φ''/φ' reduces to

$$\frac{\varphi''}{\varphi'}=\frac{[a_{23}-a_{21}(a_{43}/a_{41})]-[a_{33}-a_{31}(a_{43}/a_{41})]}{[a_{23}-a_{21}(a_{43}/a_{41})]+[a_{33}-a_{31}(a_{43}/a_{41})]}. \quad (24)$$

Re-evaluating a_{ij}, it can be shown that $[a_{23}-a_{21}(a_{43}/a_{41})]=0$ and $[a_{33}-a_{31}(a_{43}/a_{41})]\neq 0$, thus $\varphi''/\varphi'=-1$ and the transmission is zero.[7]

(*b*) At the critical angle θ_2 for the shear wave $l_\omega=1$ and $n_\omega=0$. We find for this case that the first and third rows in Eq. (11) are zeros. Consequently

$$\frac{\varphi''}{\varphi'}=\frac{(a_{22}/a_{32})-1}{(a_{22}/a_{32})+1} \quad (25)$$

which is equal to -1 only if $a_{22}=0$ and $a_{23}\neq 0$. Thus we find no transmission at θ_2 if the relationship

$$\cosh n_\Delta d=2l_\Delta(c_\omega/c_\Delta) \quad (26)$$

is satisfied, where n_Δ and l_Δ are the direction cosines corresponding to θ_2.

[7] W. C. Schneider and C. J. Burton, J. App. Phys. **20**, 48–58 (1949).

APPENDIX

Matrix elements for Eq. (11) are given with the following substitution.

$$\frac{p}{h}=c_\Delta \qquad n_\Delta=\cos\theta_\Delta$$

$$\frac{p}{k}=c_\omega \qquad n_\omega=\cos\theta_\omega$$

$$G=\rho c_\omega^2 \qquad l_\Delta=\sin\theta_\Delta$$

$$(\lambda+2Gn_\Delta^2)=\rho c_\Delta^2\cos 2\theta_\omega \qquad l_\omega=\sin\theta_\omega$$

$$D_1=-2\rho c_\Delta c_\omega[c_\omega\sin\theta_\Delta\sin 2\theta_\omega+c_\Delta\cos\theta_\omega\cos 2\theta_\omega]$$

$$D_2=[c_\Delta\cos\theta_\Delta\cos 2\theta_\omega+c_\omega\sin\theta_\omega\sin 2\theta_\Delta]$$

$$a_{11}=-\frac{2\rho c_\Delta c_\omega}{D_1}[c_\omega\sin\theta_\Delta\sin 2\theta_\omega\cos P+c_\Delta\cos\theta_\omega\cos 2\theta_\omega\cos Q]$$

$$a_{13}=-\frac{2c_\Delta c_\omega\sin\theta_\Delta\cos\theta_\omega}{D_1}[\cos P-\cos Q]$$

$$a_{21}=-\frac{2ic_\Delta c_\omega\rho}{D_1}[c_\omega\cos\theta_\Delta\sin 2\theta_\omega\sin P-c_\Delta\sin\theta_\omega\cos 2\theta_\omega\sin Q]$$

$$a_{23}=-\frac{2ic_\Delta c_\omega}{D_1}[\cos\theta_\Delta\cos\theta_\omega\sin P+\sin\theta_\Delta\sin\theta_\omega\sin Q]$$

$$a_{31}=-\frac{2c_\Delta^2c_\omega^2\rho^2}{D_1}\sin 2\theta_\omega\cos 2\theta_\omega[\cos P-\cos Q]$$

$$a_{33}=-\frac{2c_\Delta c_\omega\rho}{D_1}[c_\Delta\cos\theta_\omega\cos 2\theta_\omega\cos P+c_\omega\sin\theta_\Delta\sin 2\theta_\omega\cos Q]$$

$$a_{41}=\frac{-i\rho}{D_1}[c_\omega^2\sin 2\theta_\Delta\sin 2\theta_\omega\sin P+c_\Delta^2\cos^2 2\theta_\omega\sin Q]$$

$$a_{43}=-\frac{i}{D_1}[c_\omega\cos\theta_\omega\sin 2\theta_\Delta\sin P-c_\Delta\sin\theta_\Delta\cos 2\theta_\omega\sin Q]$$

$$a_{12}=\frac{-i}{D_2}[c_\Delta\sin\theta_\Delta\cos 2\theta_\omega\sin P+c_\omega\cos\theta_\omega\sin 2\theta_\Delta\sin Q]$$

$$a_{14}=-\frac{2ic_\Delta c_\omega}{D_2}[\sin\theta_\Delta\sin\theta_\omega\sin P+\cos\theta_\Delta\cos\theta_\omega\sin Q]$$

$$a_{22}=\frac{1}{D_2}[c_\Delta\cos\theta_\Delta\cos 2\theta_\omega\cos P+c_\omega\sin\theta_\omega\sin 2\theta_\Delta\cos Q]$$

$$a_{24}=\frac{2c_\Delta c_\omega\sin\theta_\omega\cos\theta_\Delta}{D_2}[\cos P-\cos Q]$$

$$a_{32}=\frac{i\rho}{D_2}[c_\Delta^2\cos^2 2\theta_\omega\sin P+c_\omega^2\sin 2\theta_\Delta\sin 2\theta_\omega\sin Q]$$

$$a_{34}=\frac{2i\rho c_\Delta c_\omega}{D_2}[c_\Delta\sin\theta_\omega\cos 2\theta_\omega\sin P-c_\omega\cos\theta_\Delta\sin 2\theta_\omega\sin Q]$$

$$a_{42}=\frac{\sin 2\theta_\Delta\sin 2\theta_\omega}{2D_2}[\cos P-\cos Q]$$

$$a_{44}=\frac{1}{D_2}[c_\omega\sin\theta_\omega\sin 2\theta_\Delta\cos P+c_\Delta\cos\theta_\Delta\cos 2\theta_\omega\cos Q].$$

The Theory of N Coupled Parallel Antennas*

RONOLD KING
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received July 21, 1949)

The integral-equation theory of coupled antennas developed by King and Harrison, Tai and Bouwkamp for two- and three-coupled antennas is generalized to any number of units symmetrically arranged around a circle. The case of four antennas at the corners of a square is discussed in detail. Application to cage and corner-reflector antennas is indicated.

The analysis of N parallel antennas arranged in line is carried out to determine the driving voltages required to maintain specified currents. It is shown that the inverse problem cannot be carried out in general except as a problem in simultaneous integral equations, but that the special conditions obtaining for half-wave dipoles permit an approximate analysis for the currents with given voltages. This is carried through.

1. GENERAL THEORY

THE conventional method of calculating the mutual impedance of two antennas[1–4] is based on assumed sinusoidally distributed currents on the antennas. A more rigorous method due to King and Harrison[5] and, independently, Bouwkamp,[6] solves the simultaneous integral equations for the currents in two coupled antennas and defines the self- and mutual impedances of the antennas as the coefficients of the currents in the simultaneous voltage equations. This method was improved by C. T. Tai[7] and generalized from two antennas to three arranged at the vertices of an equilateral triangle. Tai extended the analysis to N antennas symmetrically arranged in a circle only in certain special cases. This paper removes the limitations unnecessarily imposed by Tai.

The arrays of antennas to be discussed consist of N geometrically identical units each of half-length h and radius a that satisfy the inequalities, $\beta_0 a = 2\pi a/\lambda_0 \ll 1$, $a \ll h$. The several units may be center-driven by arbitary voltages or center-loaded by different impedances provided the assumed geometrical identity of the units is not significantly disturbed by the driving or load connections or the structure of generators or impedors. In practice, this means either that the driving and load connections are all alike—as when similar open-wire lines are used between all antennas and the generators or loads—or that generators and loads or the cross-sectional dimensions of transmission lines are so small in physical size that generators and loads are equivalent to dimensionless, lumped elements.

The discussion is limited to arrays of *parallel* antennas so that all currents that contribute significantly to the far-zone field are co-directional and, hence, maintain a vector potential that has no component perpendicular to the antennas that are assumed to be parallel to the z-axis of a coordinate system. That is, $\mathbf{A} = \hat{z}\mathbf{A}_z$.

Each antenna is required to have geometrical symmetry with respect to its center, and it must be center-driven or center-loaded. Furthermore, all elements have their centers in the xy-plane, i.e., they are not staggered. These conditions are imposed in order that the currents and vector potentials be even functions of z, the charges and scalar potentials odd functions of z. Thus, with complex quantities set in bold face,

$$\mathbf{I}_z(-z) = \mathbf{I}_z(z); \quad \mathbf{q}(-z) = -\mathbf{q}(z), \tag{1}$$

$$\mathbf{A}_z(-z) = \mathbf{A}_z(z); \quad \boldsymbol{\phi}(-z) = -\boldsymbol{\phi}(z). \tag{2}$$

Let it be assumed at the outset that all generators are "slice" generators, i.e., discontinuities in scalar potential, and all load impedances or tuning reactances lumped or geometrically dimensionless. These assumptions eliminate complications arising from transmission-line end-effects and coupling effects that are necessarily characteristic of particular driving and coupling circuits. Since such junction-effects are localized, they do not alter the formulation of the general problems of coupled antennas, and account may be taken of them later using appropriately designed corrective networks of lumped elements.[8,9] Accordingly, let the driving

$$C_a(h,z) \equiv \int_0^h \cos\beta_0 z' \left[\frac{\exp(-j\beta_0 R_1)}{R_1} + \frac{\exp(-j\beta_0 R_2)}{R_2}\right] dz'$$

$$S_a(h,z) \equiv \int_0^h \sin\beta_0 z' \left[\frac{\exp(-j\beta_0 R_1)}{R_1} + \frac{\exp(-j\beta_0 R_2)}{R_2}\right] dz'$$

$$R_1 \equiv ((z-z')^2+a^2)^{\frac{1}{2}}; \quad R_2 \equiv ((z+z')^2+a^2)^{\frac{1}{2}}.$$

* The research reported in this document was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5-ori-76, T. O. 1.

[1] P. S. Carter, "Circuit relations in radiating systems and applications to antenna problems," Proc. I.R.E. **20**, 1004 (1932).

[2] G. H. Brown, "Directional antennas," Proc. I.R.E. **25**, 78 (1937).

[3] B. Starnecki and E. Fitch, "Mutual impedance of two center-driven parallel aerials," S.R.D.E. Report No. 1023.

[4] K. J. Affanasiev, "Simplifications in the consideration of mutual effects between half-wave dipoles in collinear and parallel orientations," Proc. I.R.E. **34**, 635–638 (1946).

[5] R. King and C. W. Harrison, Jr., "Mutual and self-impedance for coupled antennas," J. App. Phys. **15**, 481–495 (1944).

[6] C. J. Bouwkamp, "On the theory of coupled antennae," Phillips Res. Rep. **3**, 213–226 (1948).

[7] C. T. Tai, "The theory of coupled antennas and its applications," Cruft Laboratory Technical Report No. 12 (March, 1948); "Coupled antennas," Proc. I.R.E. **36**, 487–500 (1948).

[8] R. King, "Theory of antennas driven from two-wire line," Cruft Laboratory Technical Report No. 41 (June 1948). "Antennas and two-wire lines, Part I, J. App. Phys. **20**, 832 (1949).

[9] R. King and K. Tomiyasu, "Terminal impedance and generalized two-wire line theory," Cruft Laboratory Technical Report No. 74 (April 15, 1949); Proc. I.R.E. **37**, 1134 (1949).

voltage at the center of a typical antenna k be defined as follows:

$$\mathbf{V}_{k0}=\lim_{\delta\to 0}\left[\phi_k(\delta)-\phi_k(-\delta)\right]. \quad (3)$$

If the several conditions which have been imposed thus far are satisfied, the vector potential, $\mathbf{A}_z(z)$, on the surface of each of the N antennas satisfies the same one-dimensional wave equation as when the antenna is isolated. For example, on the surface of the kth antenna the equation is:

$$\partial^2\mathbf{A}_{kz}(z)/\partial z^2+\beta_0^2\mathbf{A}_{kz}(z)=j(\beta_0^2/\omega)\mathbf{z}^i\mathbf{I}_{kz}(z)\doteq 0. \quad (4)$$

For simplicity, the term which has the internal impedance per unit length, $\mathbf{z}^i$, as a factor is set equal to zero. For good conductors its effect is negligibly small. If desired, the appropriate terms are easily added to the final solution. The general solution of the homogeneous equation for the range $0\leqq z\leqq h$ is[10]

$$\mathbf{A}_{kz}(z)=-(j/v_0)[\mathbf{C}_k\cos\beta_0 z+\tfrac{1}{2}\mathbf{V}_{k0}\sin\beta_0 z]. \quad (5)$$

The arbitrary constant $\mathbf{C}_k$ must be determined from the boundary condition at $z=h$.

The integral for the vector potential at a point z on the surface of antenna k is as follows

$$\mathbf{A}_{kz}(z)=\frac{1}{4\pi\nu_0}\int_{-h}^{h}\sum_{i=1}^{N}\mathbf{I}_i(z_i')\frac{\exp(-j\beta_0 R_{ki})}{R_{ki}}dz_i', \quad (6a)$$

where

$$R_{ki}=((z_k-z_i')^2+b_{ki}^2)^{\frac{1}{2}};\quad R_{kk}=((z_k-z_k')^2+a^2)^{\frac{1}{2}}. \quad (6b)$$

Subscripts are used on the coordinate z where necessary to indicate on which antenna the distance from the xy-plane is measured. Note that the primed coordinates locate an element of integration on the axis of an antenna, the unprimed coordinates an element on the surface; b_{ki} is the distance between the axes of antenna i and antenna k.

The substitution of (6a) in (5) gives one of N simultaneous integral equations for the N currents in the N coupled elements of the array. As usual $\zeta_0=v_0/\nu_0\doteq 376.7$ ohms.

$$\int_{-h}^{h}\sum_{i=1}^{N}\mathbf{I}_{iz}(z_i')\frac{\exp(-j\beta_0 R_{ki})}{R_{ki}}dz_i'=\frac{-j4\pi}{\zeta_0}[\mathbf{C}_k\cos\beta_0 z+\tfrac{1}{2}\mathbf{V}_{k0}\sin\beta_0 z];\qquad k=1, 2, \cdots N. \quad (7)$$

Since the solution of N simultaneous integral equations of this type has not been accomplished even for $N=2$, the only cases for which a solution is available are those in which all the N equations are alike, so that they reduce to only a single equation. Evidently, such a reduction involves restrictions in addition to those already imposed. Mathematically it implies that the current $\mathbf{I}_{kz}(z')$, which is included in the sum in (7), can be factored out in front of the sign of summation and the remaining sum is a constant that does not vary with the numerical value of k. That is, the desired equation must have the following form:

$$\int_{-h}^{h}\mathbf{I}_{kz}(z_k')\mathbf{K}_\Sigma(z, z')dz_k'=\frac{-j4\pi}{\zeta_0}[\mathbf{C}_k\cos\beta_0 z+\tfrac{1}{2}\mathbf{V}_{k0}\sin\beta_0 z], \quad (8a)$$

where

$$\mathbf{K}_\Sigma(z, z')=\sum_{i=1}^{N}\left[\frac{\mathbf{I}_{iz}(z_i')}{\mathbf{I}_{ik}(z_k')}\frac{\exp(-j\beta_0 R_{ik}}{R_{ik}}\right]$$

is independent of k, (8b)

and

$$R_{ik}=((z_i-z_k')+b_{ik}^2)^{\frac{1}{2}};\quad R_{kk}=((z_k-z_k')^2+a^2)^{\frac{1}{2}}. \quad (8c)$$

Evidently, this is not true in general. It is necessary that both the ratio of currents and the function of R_{ik} be independent of k.

Physically, the reduction of the N simultaneous integral equations to a single equation necessarily means that the N units are electrically and geometrically indistinguishable. These conditions certainly require the antennas to be located at the corners of an N-sided equilateral polygon. If this is true

$$\sum_{i=1}^{N}\exp(-j\beta_0 R_{ik})/R_{ik}$$

is obviously independent of k. The simplest way to make the current ratio in (8b) independent of k is to require all currents to be equal in magnitude and in phase, a condition achieved by identical driving voltages on all units. However, this condition is severer than required. Another, somewhat more general possibility is to have the current in each unit equal in magnitude but differing in phase from the next unit around the polygon by a fixed angle $m\theta$. That is,

$$\mathbf{I}_{iz}(z')=\mathbf{I}_{i-1,z}(z')p^m;\quad m=0, 1, 2, \cdots N-1, \quad (9a)$$

where

$$p=e^{j\theta};\quad \theta=2\pi/N. \quad (9b)$$

Or, more generally,

$$\mathbf{I}_{iz}(z')=\mathbf{I}_{kz}(z')\mathbf{p}^{(i-k)m}. \quad (10)$$

With (9a) substituted in (7), this reduces to the desired form (8a) with

$$\mathbf{K}_\Sigma(z, z')=\sum_{i=1}^{N}\mathbf{p}^{(i-k)m}\frac{\exp(-j\beta_0 R_{ik})}{R_{ik}}. \quad (11)$$

[10] R. King and D. Middleton, "The cylindrical antenna; current and impedance," Quart. App. Math. 3, 302–335 (1946).

Note that the variable of integration is changed from z_i' to z_k'. Obviously (8c) and (6b) are expressions for the *same* distances. It is readily shown that (11) is independent of k by adding an arbitrary integer q to k. Since the *order* of summation makes no difference, the same integer q may be added to i to give $i'=i+q$. The limits of summation of i' are the same as for i. Thus,

$$\mathbf{K}_\Sigma(z, z') = \sum_{i'=1}^{N} \mathbf{p}^{(i'-k-q)} \frac{\exp(-j\beta_i R_{i'k+q})}{R_{i'k+q}}. \quad (12)$$

Since $i'-k-q=i-q$, and geometrical symmetry makes

$$R_{ik} = R_{i+q,\,k+q}, \quad (13)$$

it follows that (12) is identically the same as (11). Since $\mathbf{K}_\Sigma(z, z')$ is independent of k, it is convenient to choose $k=1$ and define the kernel $\mathbf{K}_\Sigma(z, z')$ as follows:

$$\mathbf{K}_{\Sigma m}(z, z') = \sum_{i=1}^{N} \mathbf{p}^{(i-l)m} \frac{\exp(-j\beta_0 R_{il})}{R_{il}}. \quad (14)$$

Currents of equal magnitude and with a progressive, constant phase difference around the symmetrically located units are possible only when the driving voltages are equal in magnitude and have the same phase relations as the currents. That is

$$\mathbf{V}_{i0} = \mathbf{V}_{k0}\mathbf{p}^{(i-k)m}. \quad (15)$$

Since the constant $\mathbf{C}_k$ in (8a) has $\mathbf{V}_{k0}$ as a factor when it is evaluated, it follows that multiplication of (8a) by $\mathbf{p}^{(i-k)m}$ on both sides changes it to an equation for antenna i instead of for antenna k.

The integral equation for the current in the upper half of each antenna in the symmetrical array is:

$$\int_{-h}^{h} \mathbf{I}(z')\mathbf{K}_{\Sigma m}(z, z')dz' = \frac{-j4\pi}{\zeta_0}[\mathbf{C}\cos\beta_0 z + \tfrac{1}{2}\mathbf{V}_0 \sin\beta_0 z], \quad (16)$$

with $\mathbf{K}_{\Sigma m}(z, z')$ given in (14). This equation is formally exactly like that for two coupled antennas.[7] However, the kernel is more complicated since it involves N terms instead of two. The solution of (16) evidently is the same as for two antennas with $\mathbf{K}_{\Sigma m}(z, z')$ substituted for $\mathbf{K}_s(z, z')$. By mere changes in notation an expression for the current is obtained with an expansion parameter defined in terms of $\mathbf{K}_{\Sigma m}(z, z')$ instead of $\mathbf{K}_s(z, z')$. The current so obtained applies to any one of the N units, say antenna 1. Currents in the other units are obtained by multiplying this current by $\mathbf{p}^{m(i-1)}$ with $i=2, 3, \cdots N$.

It has been shown that the general problem of N-coupled, parallel, identical, non-staggered antennas reduces to a single integral equation that is like the equation for an isolated antenna provided the antennas are all located at the corners of a regular polygon and provided they are individually excited by generators or transmission lines that maintain voltages across their terminals that are equal in magnitude and vary progressively in phase by a constant angle from unit to unit around the polygon. For an array of N antennas the constant phase angle may have any one of the values $m\theta$ where $\theta=2\pi/N$ and m is an integer less than N. Since $m=0$ is a possible value, it is seen that there are N available and different angles for each of which a single integral equation governs all currents. These N possible values of m are called *phase sequences*, as is customary in the well-known method of symmetrical components. They may be defined as follows:

Zero phase sequence:

$$m=0: \quad \mathbf{V}_{10}=\mathbf{V}_{20}=\mathbf{V}_{30}=\cdots=\mathbf{V}_{N0}=\mathbf{V}^{(0)}; \quad \mathbf{I}_{10}=\mathbf{I}_{20}=\mathbf{I}_{30}=\cdots=\mathbf{I}_{N0}=\mathbf{I}^{(0)}. \quad (17a)$$

$2\pi/N$ phase sequence:

$$m=1: \quad \mathbf{V}_{10}=\mathbf{V}^{(1)};\ \mathbf{V}_{20}=\mathbf{p}\mathbf{V}^{(1)};\ \mathbf{V}_{30}=\mathbf{p}^2\mathbf{V}^{(1)};\ \cdots;\ \mathbf{V}_{N0}=\mathbf{p}^{N-1}\mathbf{V}^{(1)}; \quad \mathbf{I}_{10}=\mathbf{I}^{(1)};\ \mathbf{I}_{20}=\mathbf{p}\mathbf{I}^{(1)};\ \mathbf{I}_{30}=\mathbf{p}^2\mathbf{I}^{(1)};\ \cdots;\ \mathbf{I}_{N0}=\mathbf{p}^{N-1}\mathbf{I}^{(1)}. \quad (17b)$$

$4\pi/N$ phase sequence:

$$m=2: \quad \mathbf{V}_{10}=\mathbf{V}^{(2)};\ \mathbf{V}_{20}=\mathbf{p}^2\mathbf{V}^{(2)};\ \mathbf{V}_{30}=\mathbf{p}^4\mathbf{V}^{(2)};\ \cdots;\ \mathbf{V}_{N0}=\mathbf{p}^{2N-2}\mathbf{V}^{(2)}; \quad \mathbf{I}_{10}=\mathbf{I}^{(2)};\ I_{20}=\mathbf{p}^2\mathbf{I}^{(2)};\ I_{30}=\mathbf{p}^4\mathbf{I}^{(2)};\ \cdots;\ \mathbf{I}_{N0}=\mathbf{p}^{2N-2}\mathbf{I}^{(2)}. \quad (17c)$$

$2m\pi/N$ phase sequence:

$$m=m: \quad \mathbf{V}_{10}=\mathbf{V}^{(m)};\ \mathbf{V}_{20}=\mathbf{p}^m\mathbf{V}^{(m)};\ \mathbf{V}_{30}=\mathbf{p}^{2m}\mathbf{V}^{(m)};\ \cdots;\ \mathbf{V}_{N0}=\mathbf{p}^{mN-m}\mathbf{V}^{(m)}; \quad \mathbf{I}_{10}=\mathbf{I}^{(m)};\ \mathbf{I}_{20}=\mathbf{p}^m\mathbf{I}^{(m)};\ \mathbf{I}_{30}=\mathbf{p}^{2m}I^{(m)};\ \cdots;\ \mathbf{I}_{N0}=\mathbf{p}^{mN-m}\mathbf{I}^{(m)}. \quad (17d)$$

$2(N-1)\pi/N$ phase sequence:

$$m=N-1: \quad \mathbf{V}_{10}=\mathbf{V}^{(N-1)};\ \mathbf{V}_{20}=\mathbf{p}^{N-1}\mathbf{V}^{(N-1)};\ \mathbf{V}_{30}=\mathbf{p}^{2N-2}\mathbf{V}^{(N-1)};\ \cdots;\ \mathbf{V}_{N0}=\mathbf{p}^{N^2-2N+1}\mathbf{V}^{(N-1)}; \quad \mathbf{I}_{10}=\mathbf{I}^{(N-1)};\ \mathbf{I}_{20}=\mathbf{p}^{N-1}\mathbf{I}^{(N-1)};\ \mathbf{I}_{30}=\mathbf{p}^{2N}\mathbf{I}^{(N-1)};\ \cdots;\ \mathbf{I}_{N0}=\mathbf{p}^{N^2-2N+1}\mathbf{I}^{(N-1)}. \quad (17e)$$

Note that superscripts on $\mathbf{p}$ are powers, superscripts on $\mathbf{V}$ and $\mathbf{I}$ are written in parentheses and denote phase-sequence numbers. Note also that

$$\mathbf{p}^N = e^{j2\pi} = 1;\ \mathbf{p}^{N/2} = e^{j\pi} = -1; \quad \mathbf{p}^{N/4} = e^{j\pi/2} = j;\ \mathbf{p}^{3N/4} = e^{j3\pi/2} = -j. \quad (18)$$

The current $\mathbf{I}^{(m)}$ in each antenna can be determined independently for each phase sequence with an arbitrary value of driving voltage $\mathbf{V}^{(m)}$ using (16). If each antenna is imagined to be driven by all voltages in all N

phase sequences simultaneously, with the N voltages in each unit in series, the following driving voltages apply:

$$\mathbf{V}_{10}=\mathbf{V}^{(0)}+\mathbf{V}^{(1)}+\mathbf{V}^{(2)}+\cdots+\mathbf{V}^{(N-1)}=\sum_{i=0}^{N-1}\mathbf{V}^{(i)} \tag{19a}$$

$$\mathbf{V}_{20}=\mathbf{V}^{(0)}+\mathbf{p}\mathbf{V}^{(1)}+\mathbf{p}^2\mathbf{V}^{(2)}+\cdots+\mathbf{p}^{N-1}\mathbf{V}^{(N-1)}=\sum_{i=0}^{N-1}\mathbf{p}^i\mathbf{V}^{(i)} \tag{19b}$$

$$\mathbf{V}_{30}=\mathbf{V}^{(0)}+\mathbf{p}^2\mathbf{V}^{(1)}+\mathbf{p}^4\mathbf{V}^{(2)}+\cdots+\mathbf{p}^{2N-2}\mathbf{V}^{(N-1)}=\sum_{i=0}^{N-1}\mathbf{p}^{2i}\mathbf{V}^{(i)} \tag{19c}$$

$$\vdots$$

$$\mathbf{V}_{N0}=\mathbf{V}^{(0)}+\mathbf{p}^{N-1}\mathbf{V}^{(1)}+\mathbf{p}^{2N-2}\mathbf{V}^{(2)}+\cdots+\mathbf{p}^{N-2N+1}\mathbf{V}^{(N-1)}=\sum_{i=0}^{N-1}\mathbf{p}^{Ni-i}\mathbf{V}^{(i)}. \tag{19d}$$

Similarly the resultant currents in the antennas are:

$$\mathbf{I}_{10}=\sum_{i=0}^{N-1}\mathbf{I}^{(i)} \tag{20a}$$

$$\mathbf{I}_{20}=\sum_{i=0}^{N-1}\mathbf{p}^i\mathbf{I}^{(i)} \tag{20b}$$

$$\vdots$$

$$\mathbf{I}_{N0}=\sum_{i=0}^{N-1}\mathbf{p}^{Ni-i}\mathbf{I}^{(i)}. \tag{20c}$$

Since the N voltages $\mathbf{V}^{(i)}$, $i=0, 1, 2, \cdots N-1$, are arbitrary, it is possible to determine a set of N such voltages which satisfy (19a–d) for all possible values of $\mathbf{V}_{10}, \mathbf{V}_{20}, \cdots \mathbf{V}_{N0}$. That is, the general problem of solving for the N currents in N coupled antennas located at the corners of a regular polygon may be carried out by first solving for the N independent currents due to voltages chosen to satisfy the conditions of the N phase sequences. An appropriate superposition of these N solutions yields the currents and impedances in all units due to an arbitrary set of driving voltages.

As a simple illustration, let the general theory be applied to two coupled antennas. In this case $N=2$ and m is restricted to the values 0 and 1. That is, there are two phase sequences, the zero phase sequence previously called the symmetrical case, and the π phase-sequence previously called the antisymmetrical case.

Equations corresponding to (19) and (20) are

$$\mathbf{V}_{10}=\mathbf{V}^{(0)}+\mathbf{V}^{(1)}=\mathbf{V}^s+\mathbf{V}^a, \tag{21a}$$

$$\mathbf{V}_{20}=\mathbf{V}^{(0)}+\mathbf{p}\mathbf{V}^{(1)}=\mathbf{V}^{(0)}-\mathbf{V}^{(1)}=\mathbf{V}^s-\mathbf{V}^a. \tag{21b}$$

The corresponding kernels for the integral equation are obtained from (14) with $m=0$ and 1 to be

$$\mathbf{K}_{\Sigma 0}(z, z')\equiv\mathbf{K}_s(z, z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}+\frac{\exp(-j\beta_0 R_{12})}{R_{12}} \tag{22a}$$

$$\mathbf{K}_{\Sigma 1}(z, z')\equiv\mathbf{K}_a(z, z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}+\frac{\exp(-j\beta_0 R_{12})}{R_{12}}. \tag{22b}$$

These results are the same as previously obtained.[7] The theory for N antennas in terms of N phase sequences is a simple generalization of the theory developed by Tai for two and three antennas with two and three phase sequences. Its application to four antennas follows.

2. FOUR ANTENNAS AT CORNERS OF A SQUARE

The integral equation for the current in the upper half of each of four identical, parallel, non-staggered antennas placed at the corners of a square of side $b_{12}=b$ and diagonal $b_{13}=b\sqrt{2}$ is the same as (1.16). The kernel in this case is obtained from (1.15) with $n=4$. It is

$$\mathbf{K}_{\Sigma m}(z, z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}+\mathbf{p}^m\frac{\exp(-j\beta_0 R_{12})}{R_{12}}+\mathbf{p}^{2m}\frac{\exp(-j\beta_0 R_{13})}{R_{13}}+\mathbf{p}^{3m}\frac{\exp(-j\beta_0 R_{14})}{R_{14}} \tag{1}$$

where

$$\mathbf{p}=e^{j\theta};\ \theta=\pi/2 \tag{2a}$$

so that

$$\mathbf{p}=j;\ \mathbf{p}^2=-1;\ \mathbf{p}^3=-j;\ \mathbf{p}^4=1. \tag{2b}$$

With $n=4$, it follows that

$$m=0, 1, 2, 3. \tag{4}$$

The four phase sequences are illustrated schematically in Fig. 1. They include the zero-phase sequence ($m=0$) with all antennas in phase; the $\pi/4$ or positive phase sequence ($m=1$) with antenna $k+1$ leading antenna k by $\pi/4$, $k=1, 2, 3, 4$; the π-phase sequence ($m=2$) with currents alternating in direction around the square; and the $3\pi/4$, $-\pi/4$, or negative phase sequence with antenna $k+1$ leading antenna k by $3\pi/4$, $k=1, 2, 3, 4$.

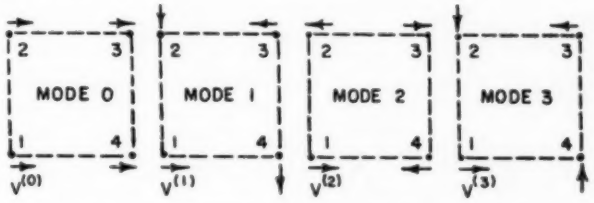


FIG. 1. The four sequences of voltages occurring in four coupled antennas.

The four kernels are

$$m=0:\quad \mathbf{K}_{\Sigma 0}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}+\frac{\exp(-j\beta_0 R_{12})}{R_{12}}+\frac{\exp(-j\beta_0 R_{13})}{R_{13}}+\frac{\exp(-j\beta_0 R_{14})}{R_{14}} \tag{5a}$$

$$m=1:\quad \mathbf{K}_{\Sigma 1}(z,z')=\frac{\exp-(j\beta_0 R_{11})}{R_{11}}+j\frac{\exp(-j\beta_0 R_{12})}{R_{12}}-\frac{\exp(-j\beta_0 R_{13})}{R_{13}}-j\frac{\exp(-j\beta_0 R_{14})}{R_{14}} \tag{5b}$$

$$m=2:\quad \mathbf{K}_{\Sigma 2}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}-\frac{\exp(-j\beta_0 R_{12})}{R_{12}}+\frac{\exp(-j\beta_0 R_{13})}{R_{13}}-\frac{\exp(-j\beta_0 R_{14})}{R_{14}} \tag{5c}$$

$$m=3:\quad \mathbf{K}_{\Sigma 3}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}-\frac{\exp(-j\beta_0 R_{12})}{R_{12}}-\frac{\exp(-j\beta_0 R_{13})}{R_{13}}+j\frac{\exp(-j\beta_0 R_{14})}{R_{14}}. \tag{5d}$$

These expressions can be simplified if use is made of the geometrical condition $R_{12}=R_{14}$ with which (5) becomes

$$m=0:\quad \mathbf{K}_{\Sigma 0}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}+2\frac{\exp(-j\beta_0 R_{12})}{R_{12}}+\frac{\exp(-j\beta_0 R_{13})}{R_{13}} \tag{6a}$$

$$m=1,3:\quad \mathbf{K}_{\Sigma 1}(z,z')=\mathbf{K}_{\Sigma 3}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}-\frac{\exp(-j\beta_0 R_{13})}{R_{13}} \tag{6b}$$

$$m=2:\quad \mathbf{K}_{\Sigma 2}(z,z')=\frac{\exp(-j\beta_0 R_{11})}{R_{11}}-2\frac{\exp(-j\beta_0 R_{12})}{R_{12}}+\frac{\exp(-j\beta_0 R_{13})}{R_{13}}. \tag{6c}$$

If these three kernels are substituted in

$$\mathbf{\Psi}_{\Sigma m}(z)\sin\beta_0(h-|z|)=\int_{-h}^{h}\mathbf{K}_{\Sigma m}(z,z')\sin\beta_0(h-z')dz', \tag{7a}$$

the corresponding expansion parameters may be evaluated from

$$\Psi_{\Sigma m}=\left\{\begin{array}{ll}|\mathbf{\Psi}_{\Sigma m}(0)|; & \beta_0 h\leqq\pi/2\\ |\mathbf{\Psi}_{\Sigma m}(h-\lambda_0/4)|; & \beta_0 h\geqq\pi/2\end{array}\right\} \tag{7b}$$

with

$$\mathbf{\Psi}_{\Sigma 0}(z)\sin\beta_0(h-z)=[C_a(h,z)+2C_b(h,z)+C_c(h,z)]\sin\beta_0 h+[S_a(h,z)+2S_b(h,z)+S_c(h,z)]\cos\beta_0 h. \tag{8a}$$

$$\mathbf{I}_{\Sigma 1}(z)\sin\beta_0(h-z)=[C_a(h,z)-C_c(h,z)]\sin\beta_0 h+[S_a(h,z)-S_c(h,z)]\cos\beta_0 h \tag{8b}$$

$$\mathbf{I}_{\Sigma 2}(z)\sin\beta_0(h-z)=[C_a(h,z)-2C_b(h,z)+C_c(h,z)]\sin\beta_0 h+[S_a(h,z)-2S_b(h,z)+S_c(h,z)]\cos\beta_0 h. \tag{8c}$$

The function $C_c(h,z)$ and $S_c(h,z)$ are like the corresponding functions with subscript a defined by Tai[7] with c substituted for a, where $c=b_{13}=b\sqrt{2}$.

With the expansion parameters for the four phase sequences evaluated, the corresponding currents and impedances may be determined just as for two and three antennas.

The general case in which each of the four antennas is driven by an arbitrary voltage may be solved by superposition using the following relations:

$$\mathbf{V}_{10}=\mathbf{V}^{(0)}+\mathbf{V}^{(1)}+\mathbf{V}^{(2)}+\mathbf{V}^{(3)} \tag{9a}$$

$$\mathbf{V}_{20}=\mathbf{V}^{(0)}+j\mathbf{V}^{(1)}-\mathbf{V}^{(2)}-j\mathbf{V}^{(3)} \tag{9b}$$

$$\mathbf{V}_{30}=\mathbf{V}^{(0)}-\mathbf{V}^{(1)}+\mathbf{V}^{(2)}-\mathbf{V}^{(3)} \tag{9c}$$

$$\mathbf{V}_{40}=\mathbf{V}^{(0)}-j\mathbf{V}^{(1)}-\mathbf{V}^{(2)}+j\mathbf{V}^{(3)}. \tag{9d}$$

Alternatively, solving for the phase-sequence voltages,

$$\mathbf{V}^{(0)}=\tfrac{1}{4}(\mathbf{V}_{10}+\mathbf{V}_{20}+\mathbf{V}_{30}+\mathbf{V}_{40}) \tag{10a}$$

$$\mathbf{V}^{(1)}=\tfrac{1}{4}(\mathbf{V}_{10}-j\mathbf{V}_{20}-\mathbf{V}_{30}+j\mathbf{V}_{40}) \tag{10b}$$

$$\mathbf{V}^{(2)}=\tfrac{1}{4}(\mathbf{V}_{10}-\mathbf{V}_{20}+\mathbf{V}_{30}-\mathbf{V}_{40}) \tag{10c}$$

$$\mathbf{V}^{(3)}=\tfrac{1}{4}(\mathbf{V}_{10}+j\mathbf{V}_{20}-\mathbf{V}_{30}-j\mathbf{V}_{40}). \tag{10d}$$

The same equations apply to the currents if $\mathbf{I}$ is substituted for $\mathbf{V}$ in (9) and (10). By setting

$$\mathbf{V}^{(0)}=\mathbf{I}^{(0)}\mathbf{Z}^{(0)};\ \mathbf{V}^{(1)}=\mathbf{I}^{(1)}\mathbf{Z}^{(1)};\ \mathbf{V}^{(2)}=\mathbf{I}^{(2)}\mathbf{Z}^{(2)};\ \mathbf{V}^{(3)}=\mathbf{I}^{(3)}\mathbf{Z}^{(3)} \tag{11}$$

in (9) and substituting (10) with $\mathbf{I}$ written for $\mathbf{V}$ in the resulting equations, the following results are obtained:

$$\mathbf{V}_{i0}=\sum_{j=1}^{4}\mathbf{I}_{j0}\mathbf{Z}_{ij};\quad i=1,2,3,4. \tag{12}$$

The evaluation of the self-impedance $\mathbf{Z}_{ii}$ and mutual impedances $\mathbf{Z}_{ij}$ is facilitated if note is taken of the following relations which are a consequence of geometrical and electrical symmetry and the reciprocal theorem:

$$\mathbf{Z}^{(1)}=\mathbf{Z}^{(3)} \tag{13a}$$

$$\mathbf{Z}_{11}=\mathbf{Z}_{22}=\mathbf{Z}_{33}=\mathbf{Z}_{44} \tag{13b}$$

$$\mathbf{Z}_{12}=\mathbf{Z}_{23}=\mathbf{Z}_{34}=\mathbf{Z}_{41}=\mathbf{Z}_{21}=\mathbf{Z}_{32}=\mathbf{Z}_{43}=\mathbf{Z}_{14} \tag{13c}$$

$$\mathbf{Z}_{13}=\mathbf{Z}_{24}=\mathbf{Z}_{31}=\mathbf{Z}_{42}. \tag{13d}$$

With (13a–d) the following impedances are obtained directly:

$$\mathbf{Z}_{11}=\tfrac{1}{4}[\mathbf{Z}^{(0)}+2\mathbf{Z}^{(1)}+\mathbf{Z}^{(2)}] \tag{14a}$$

$$\mathbf{Z}_{12}=\tfrac{1}{4}[\mathbf{Z}^{(0)}-\mathbf{Z}^{(2)}] \tag{14b}$$

$$\mathbf{Z}_{13}=\tfrac{1}{4}[\mathbf{Z}^{(0)}-2\mathbf{Z}^{(1)}+\mathbf{Z}^{(2)}]. \tag{14c}$$

Alternatively, solving for the sequence impedances,

$$\mathbf{Z}^{(0)}=\mathbf{Z}_{11}+2\mathbf{Z}_{12}+\mathbf{Z}_{13} \tag{15a}$$

$$\mathbf{Z}^{(1)}=\mathbf{Z}_{11}-\mathbf{Z}_{13} \tag{15b}$$

$$\mathbf{Z}^{(2)}=\mathbf{Z}_{11}-2\mathbf{Z}_{12}+\mathbf{Z}_{13}. \tag{15c}$$

The symmetry between the phase-sequence impedances and the self- and mutual impedances in (15a–c) and the phase-sequence kernels and the exponential terms in (6a–c) is apparent.

With (14a–c) used in (12) with (13), the currents in the four antennas due to arbitrary driving voltages may be determined.

3. CAGE ANTENNAS

Cage antennas are arrays of N closely spaced parallel antennas excited in the *zero phase sequence* so that *all currents are in phase*. The kernel for an N unit array is

$$\mathbf{K}_{\Sigma 0}(z,z')=\sum_{i=1}^{N}\frac{\exp(-j\beta_0 R_{1i})}{R_{1i}} \tag{1}$$

and the expansion function is

$$\mathbf{\Psi}_{\Sigma 0}(z)=\int_{-h}^{h} g(z,z')\mathbf{K}_{\Sigma 0}(z,z')dz', \tag{2}$$

where $g(z,z')=\sin\beta_0(h-z')/\sin\beta_0(h-z)$. This may be expressed as follows:

$$\mathbf{\Psi}_{\Sigma 0}(z)=\mathbf{\Psi}_{K1}(z)+\mathbf{\Psi}_{b12}(z)=\mathbf{\Psi}_{b13}(z)+\cdots \tag{3}$$

where

$$\mathbf{\Psi}_{K1}(z)=\int_{-h}^{h} g(z,z')\frac{\exp(-j\beta_0 R_{11})}{R_{11}}dz';$$

$$R_{11}=((z-z')^2+a^2)^{\frac{1}{2}} \tag{4a}$$

$$\mathbf{\Psi}_{b12}(z)=\int_{-h}^{h} g(z,z')\frac{\exp(-j\beta_0 R_{12})}{R_{12}}dz';$$

$$R_{12}=((z-z')^2+b_{12}^2)^{\frac{1}{2}} \tag{4b}$$

$$\mathbf{\Psi}_{b13}(z)=\int_{-h}^{h} g(z,z')\frac{\exp(-j\beta_0 R_{13})}{R_{13}}dz';$$

$$R_{13}=((z-z')^2+b_{13}^2)^{\frac{1}{2}} \tag{4c}$$

$$\mathbf{\Psi}_{b1N}(z)=\int_{-h}^{h} g(z,z')\frac{\exp(-j\beta_0 R_{1N})}{R_{1N}}dz';$$

$$R_{1N}=((z-z')^2+b_{1N}^2)^{\frac{1}{2}}. \tag{4d}$$

The distance between centers of antennas 1 and 2 is b_{12}, between antennas 1 and N it is b_{1N}.

For antennas sufficiently closely spaced so that

$$\beta_0 b_{1N}\ll 1, \tag{5}$$

$$\mathbf{\Psi}_{b1N}(z_r)=\mathbf{\Psi}_{K1}(z_r)-2\ln(b_{1N}/a), \tag{6}$$

where $z_r=0$ for $\beta_0 h\leqq\pi/2$, $z_r=h-\lambda_0/4$ for $\beta_0 h\geqq\pi$. Hence,

$$\Psi_{\Sigma 0}=|\mathbf{\Psi}_{\Sigma 0}(z_r)|=|N\mathbf{\Psi}_{K1}(z_r)-2\sum_{i=2}^{N}\ln b_{1i}/a| \tag{7a}$$

$$=|N\mathbf{\Psi}_{K1}(z_r)-2\ln(b_a b_{13}b_{14}\cdots b_{1N}/a^{N-1})|\cdot \tag{7b}$$

Using this formula for the expansion parameter, the current in and impedance for each of the N antennas forming the cage may be determined just as for a single antenna.

For antennas that are not too long, a satisfactory approximate value of the expansion parameter for a single antenna is

$$\Psi_{K1}=2[\ln 2h/a-1]. \tag{8}$$

Since the contribution to Ψ_{K1} by the imaginary part of $\mathbf{\Psi}_{K1}(z_r)$ is not great, an approximate formula for (7b) is obtained by replacing $\mathbf{\Psi}_{K1}(z_r)$ by Ψ_{K1} as given by (8). The result is

$$\Psi_{\Sigma 0}\doteq N\Psi_{K1}-2\ln(b_{12}b_{13}b_{14}\cdots b_{1N}/a^{N-1}) \tag{9a}$$

$$\doteq 2[\ln(2h)^N/ab_{12}b_{13}\cdots b_{1N}]. \tag{9b}$$

It is a simple matter to determine the radius a_e of a single antenna carrying N times the current of each unit in the cage and having $1/N$ times the impedance. It is necessary merely to set $a=a_e$ in (8), multiply (8) by N, and equate the result to (9b). Thus solving for a_e

$$a_e=(ab_{12}b_{13}b_{14}\cdots b_{1N})^{1/N}. \tag{10}$$

Therefore, a cage antenna consisting of N *closely spaced*, parallel, identical units arranged at regular intervals around a circle is *approximately* equivalent to a single antenna of much larger radius, in the sense that the total current and its axial distribution are approximately the same, and its impedance is $1/N$ that of the individual units if these are all driven separately.

4. CORNER REFLECTOR ANTENNAS

The obvious application of the general theory of N antennas to the corner reflector antenna using the theorem of images has been carried out by Tai[7] in general, since his analysis did not exclude the case of N antennas when the phases alternate.

5. PARALLEL ARRAYS WITH ALL ELEMENTS DRIVEN; BROADSIDE AND END-FIRE ARRAYS; CIRCUIT PROPERTIES[11]

The most important arrays of identical, parallel, non-staggered, center-driven antennas consist of N units

[11] The three-element broadside array was analyzed by C. W. Harrison, Jr., Proc. I.R.E. **34**, 204 (1946) using a method which did not take account of the coupled antennas in the kernel but only in a correction term. By limiting himself to the three-element broadside array, Harrison had to deal with only two simultaneous integral equations which he solved approximately.

spaced at equal intervals b along a straight line as shown in Fig. 2. In one class of arrays, the N currents are made equal in magnitude by appropriate driving voltages, and it is the phase relations which determine the different field characteristics of different arrays. For example, in the broadside array all currents are in phase; in the end-fire array the phase of the currents increases from one antenna to the next by an angle $\beta_0 b$. In another class of arrays, the magnitudes of the currents decrease in a prescribed manner from the center of the array outward in both directions.

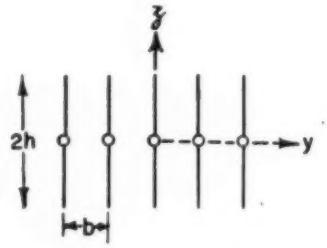


FIG. 2. Five-element parallel array.

Since there are no conditions of symmetry in an array of antennas uniformly spaced along a straight line which make it possible to reduce the N simultaneous integral equations given in (1.7) to a single equation, the conclusion is inevitable that the determination of the N currents and N impedances in terms of arbitrary driving voltages cannot be carried out with available mathematical techniques. Fortunately, an approximate solution of the important inverse problem of determining the N voltages required to maintain N currents with specified relative magnitudes and phases is possible.

The integral equations for the currents in the upper halves ($0 \leqq z \leqq h$) of the N coupled antennas are given by (1.7). They may be expressed as follows:

$$\int_{-h}^{h} \mathbf{I}_{mz}(z_m')\mathbf{K}_p(z_m, z_m')dz_m = \frac{-j4\pi}{\zeta_0}[\mathbf{C}_m \cos\beta_0 z_m + \tfrac{1}{2}\mathbf{V}_{0m} \sin\beta_0 z_m]; \quad m = 1, 2, \cdots N, \quad (1a)$$

where the kernel is

$$\mathbf{K}(z_m, z_m') = \sum_{i=1}^{N} \frac{\mathbf{I}_{iz}(z')}{\mathbf{I}_{mz}(z_m')} \frac{\exp(-j\beta_0 R_{mi})}{R_{mi}}; \quad m = 1, 2, \cdots N, \quad (1b)$$

and where

$$R_{mi} = ((z_m - z_i')^2 + b_{mi}^2)^{\frac{1}{2}}; \; R_{mm} = ((z_m - z_m') + a^2)^{\frac{1}{2}}. \quad (1c)$$

The distance between the centers of antenna m and antenna i is $b_{mi} = |m-i|\,b$, where b is the equal distance between adjacent units.

The N kernels $\mathbf{K}_p(z_m, z_m')$ are functions of the distributions of current in all antennas and of the distances R_{mi}. They are not independent. However, an approximate solution of (1a) may be obtained by assuming that the *distributions* of current as functions of z' are the same in all N units, but without thereby restricting the magnitudes or the phases of the input currents at $z=0$. It is readily agreed that this is a reasonable assumption if the antennas are identical. Indeed, it has already been shown in the analysis of two coupled antennas[7] driven either in phase or 180 degrees out of phase that the distributions of current do not differ significantly from each other or from the distribution along an isolated antenna provided $\beta_0 h$ does not greatly exceed $\pi/2$. When $\beta_0 h$ is near π the distributions differ considerably for very closely spaced antennas, but as the separation is increased, the currents become more and more nearly alike. Let it be assumed, therefore, that it is a good approximation to set

$$\frac{\mathbf{I}_{iz}(z_i')}{\mathbf{I}_{mz}(z_m')} \doteq k_{mi} \exp(j\theta_{mi}), \quad (2)$$

where k_{mi} is the specified real ratio of magnitudes and δ_{mi} the specified phase difference between the currents in antennas k and m except when the antennas are close together *and* have half-lengths h near $\lambda_0/2$.

Substitution of (2) in (1a) gives:

$$\mathbf{K}_p(z_m, z_m') = \sum_{i=1}^{N} k_{mi} \exp(j\theta_{mi}) \frac{\exp(-j\beta_0 R_{mi})}{R_{mi}}; \quad m = 1, 2, \cdots N. \quad (3)$$

The N kernels defined in (3) are not alike. They are not and cannot be made independent of m since for geometrical reasons,

$$R_{mi} \neq R_{m+q,\, i+q}, \quad (4)$$

where q is an integer for all values of m and q from 1 to N.

However, since all quantities in (3) are assumed to be known for all values of m, the N kernels can be evaluated independently. In particular, (1a) can be solved separately for $\mathbf{I}_{mz}(z)$ for each value of m. Thus, it is seen that the restrictions on the current contained in (2) have reduced the N *simultaneous* integral equations for determining N independent currents, to N *independent* integral equations for determining the N voltages $\mathbf{V}_{m0}$ that are required to maintain the N specified currents.

Since (1a) differs from the corresponding equation for two antennas only in the number of terms in the kernel, the formal solution for the current and the impedance as a function of the driving voltage may be carried out directly.[7] The expression for the input impedance of antenna m in the presence of $N-1$ parallel antennas all so driven that they have pre-assigned currents is like that for two antennas[7] with additional P

and **Q** terms each multiplied by the appropriate ratio factor and phase factor for the current. Thus,

$$(\mathbf{Z}_m)_{in}=\frac{-j\zeta_0\Psi_{pm}}{2\pi}\left\{\frac{\Psi_{pm}\cos\beta_0 h+\mathbf{F}_1(h)+\mathbf{P}_{1z}(h)}{[\Psi_{pm}+\mathbf{F}_1(0)+\mathbf{P}_{1z}(0)]\sin\beta_0 h-[\mathbf{G}_1(0)+\mathbf{Q}_{1z}(0)]\cos\beta_0 h+[\mathbf{G}_1(h)+\mathbf{Q}_{1z}(h)]}\right\}, \tag{5}$$

where

$$\Psi_{pm}=|\Psi_{pm}(0)|\,;\quad \beta_0 h\leqq\pi/2 \tag{6a}$$

$$\Psi_{pm}=|\Psi_{pm}(h-\lambda_0/4)|\,;\ \beta_0 h\geqq\pi/2 \tag{6b}$$

$$\Psi_{pm}(z)\sin\beta_0(h-z)=[C_a(h,z)+\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\times\exp(j\theta_{mi})Cb_{mi}(h,z)]\sin\beta_0 h$$

$$-[S_a(h,z)+\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_{mi})Sb_{mi}(h,z)]\cos\beta_0 h. \tag{6c}$$

Also,

$$\mathbf{P}_1(z)=\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_{mi})\mathbf{P}_{1mi}(z) \tag{7a}$$

$$\mathbf{Q}_1(z)=\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_{mi})\mathbf{Q}_{1mi}(z) \tag{7b}$$

$$\mathbf{P}_{1mi}(z)=-\int_{-h}^{h}\mathbf{F}_{1z'}\frac{\exp(-j\beta_0 R_{mi})}{R_{mi}}dz'$$

$$=-Cb_{mi}(h,z)+Eb_{mi}(h,z)\cos\beta_0 h \tag{8a}$$

$$\mathbf{Q}_{1mi}(z)=-\int_{-h}^{h}\mathbf{G}_{1z'}\frac{\exp(-j\beta_0 R_{mi})}{R_{mi}}dz'$$

$$=-Sb_{mi}(h,z)+Eb_{mi}(h,z)\sin\beta_0 h \tag{8b}$$

$$R_{mi}=((z_m-z_i')^2+b_{mi}^2)^{\frac{1}{2}}. \tag{8c}$$

$\mathbf{F}_1(0)$, $\mathbf{F}_1(h)$, $\mathbf{G}_1(0)$ and $\mathbf{G}_1(h)$ as well as $C(h,z)$, $S(h,z)$ and $E(h,z)$ are as in reference 7 with Ψ_{pm} substituted for Ψ. Note that b_{mi} is the distance between centers of antennas m and i. Note that when $N=2$, $\theta_{mi}=0$, $k_{mi}=1$, (5) reduces to the previous formula[7] for two symmetrically driven antennas. Alternatively, when $N=2$, $\theta_{mi}=\pi$, $k_{mi}=1$, (5) reduces to the formula for two antisymmetrically driven antennas.

By evaluation of the functions $\mathbf{P}_{1z}$ and $\mathbf{Q}_{1z}$ together with the expansion parameter Ψ_{pm} for *each* value of m, the input impedance of each antenna in the presence of the others may be evaluated. Note that, in general, the antennas are in identical pairs except for the central unit for N odd, so that the number of different values of these functions is $N/2$ for N even, $(N+1)/2$ for N odd. Multiplication of the impedances so determined by the appropriate, assigned, input current gives the driving voltages required to maintain these currents.

In practice, antennas in parallel arrays usually have electrical half-lengths $\beta_0 h$ that are $\pi/2$ or near $\pi/2$. Significantly, this is precisely the value of $\beta_0 h$ for which the initial assumption of identical *distributions* (but not magnitudes or phases) of current is best satisfied and for which a maximum simplification of the intricate formula (5) is possible. With $\beta_0 h=\pi/2$, (5) becomes:

$$(\mathbf{Z}_m)_{in}=\frac{j\zeta_0}{4\pi}\left\{\frac{C_a(h,h)+\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_{mi})Cb_{mi}(h,h)}{1-\frac{1}{2\Psi_{pm}}[C_a(h,0)+S_a(h,h)-E_a(h,h)+\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_m)\{Cb_{mi}(h,0)+Sb_{mi}(h,h-Eb_{mi}(h,h)\}]}\right\}. \tag{9}$$

Except for antennas that are extremely closely spaced, the expansion parameter Ψ_{pm} cannot differ greatly from the value Ψ_{K1} for each of the antennas when isolated. In most driven arrays of the types considered in this section, b is not less than $\lambda_0/4$ and frequently is $\lambda_0/2$. Since the term with $1/2\Psi_{pm}$ as a factor in the denominator of (9) necessarily is small compared with unity for antennas of small radius it may be neglected. That this is indeed a good approximation already has been demonstrated for two coupled antennas[7] with $\beta_0 h=\pi/2$. If this term is neglected, (9) reduces to

$$(\mathbf{Z}_m)_{in}=\frac{j\zeta_0}{4\pi}[C_a(h,h)+\sum_{\substack{i=1\\ i\neq m}}^{N} k_{mi}\exp(j\theta_{mi})Cb_{mi}(h,h)];$$

$$m=1,2,\cdots N. \tag{10}$$

Multiplication of (10) by $\mathbf{I}_{m0}$, noting that

$$\mathbf{I}_{i0}=\mathbf{I}_{m0}k_{mi}\exp(j\theta_{mi}), \tag{11}$$

reveals that (10) is equivalent to

$$\mathbf{V}_{10}=\mathbf{I}_{10}\mathbf{Z}_{11}+\mathbf{I}_{20}\mathbf{Z}_{12}+\mathbf{I}_{30}\mathbf{Z}_{13}+\cdots+\mathbf{I}_{N0}\mathbf{Z}_{1N},$$

$$\mathbf{V}_{20}=\mathbf{I}_{10}\mathbf{Z}_{21}+\mathbf{I}_{20}\mathbf{Z}_{22}+\mathbf{I}_{30}\mathbf{Z}_{23}+\cdots+\mathbf{I}_{N0}\mathbf{Z}_{2N},$$

$$\cdots$$

$$V_{N0}=\mathbf{I}_{10}\mathbf{Z}_{N1}+\mathbf{I}_{20}\mathbf{Z}_{N2}+\mathbf{I}_{30}\mathbf{Z}_{N3}+\cdots+\mathbf{I}_{N0}\mathbf{Z}_{NN}, \quad (12)$$

where

$$\mathbf{Z}_{11}=\mathbf{Z}_{22}=\mathbf{Z}_{33}=\cdots\mathbf{Z}_{NN}=\frac{j\zeta_0}{2\pi}C_a(h,h)=\mathbf{Z}_0 \quad (13)$$

$$\mathbf{Z}_{12}=\mathbf{Z}_{21}=\frac{j\zeta_0}{2\pi}C_{b12}(h,h) \quad (14a)$$

$$\mathbf{Z}_{13}=\mathbf{Z}_{31}=\frac{j\zeta_0}{2\pi}C_{b13}(h,h) \quad (14b)$$

$$\cdots$$

$$\mathbf{Z}_{1N}=\mathbf{Z}_{N1}=\frac{j\zeta_0}{2\pi}C_{b1N}(h,h). \quad (14c)$$

Note that the self-impedances $\mathbf{Z}_{ii}$ of the coupled antennas are all the same as the self-impedance $\mathbf{Z}_0$ of an isolated antenna. Moreover, the mutual impedances $\mathbf{Z}_{ij}$ are the same as if only antennas i and j were present. Therefore, it may be concluded that for parallel arrays in which antennas of half-length near $h=\lambda_0/4$ are coupled, the self-impedance of an isolated antenna may be used for each antenna and the mutual impedance between any pair taken to be the same as for that pair when isolated. Accordingly, the self-impedance and the mutual impedance may be obtained from first-order curves of Z_{s1},[7] for the appropriate value of $\Omega=2\ln 2h/a$.

It is significant to note that in spite of the initial *general* statement to the contrary, the Eqs. (13) may be solved for either the driving voltages given the current or the currents given the driving voltages. However, this is true only approximately and only in the special case of antennas for which $\beta_0 h=\pi/2$ and Ψ_{12} is not too small.

The simple method of analyzing N coupled antennas in terms of self-impedances of isolated antennas and mutual impedances of two coupled antennas obviously is *not valid* for antennas with half-lengths that exceed $\lambda_0/4$ considerably. Specifically, when $\beta_0 h=\pi$, (5) becomes

$$(\mathbf{Z}_m)_{in}=\frac{j\zeta_0\Psi_{pm}^2}{2\pi}\times\left\{\frac{1-(1/\Psi_{pm})[\mathbf{F}_1(h)+\mathbf{P}_1(h)]}{[\mathbf{G}_1(0)+\mathbf{Q}_{12}(0)+\mathbf{G}_1(h)+\mathbf{Q}_{12}(h)]}\right\}. \quad (15)$$

Since the term involving the most significant contributions from the coupled antennas now appear in the factor Ψ_{pm}^2 and in the terms in $\mathbf{Q}$ in the denominator, it is not possible to express $\mathbf{V}_{m0}=\mathbf{I}_{m0}(\mathbf{Z}_m)_{in}$ as a simple sum of self- and mutual terms. $(\mathbf{Z}_m)_{in}$ can be evaluated from (16) and $\mathbf{V}_{m0}$ evaluated for each antenna if the currents are known. The reverse is not possible.

For the broadside array, $k_{mi}=1$, $\theta_{mi}=0$ with $\beta_0 b=\pi$. For the end-fire array, $\theta_{mi}=\pm(m-i)\beta_0 b$ with $\beta_0 b=\pi/2$ for the unilateral end-fire, π for bilateral end-fire. The analysis in this section has demonstrated that common assumptions, such as that equal driving voltages will produce equal currents in all units of a broadside array, are erroneous except for the two-element array. The different impedances of the differently situated units require different voltages if the currents are to be the same.

Conventional methods of driving parallel arrays in which currents in the units are assumed to be respectively in phase (broadside array) or alternately 180 degrees out-of-phase (bilateral end-fire array) are shown in Fig. 3 and Fig. 4 for three-element arrays. The transmission line connecting the units in the broadside array is assumed to be *spiraled* and *not crossed-over* as shown in the figure so that it is the equivalent of a smooth line one-half wave-length long. Actually, the currents in the three antennas in Figs. 3 and 4 are neither all equal in magnitude nor all in phase (broadside) or alternately 180 degrees out-of-phase (bilateral end-fire) for two reasons. These are, (1) the input impedance of the central unit differs from that of the two outer units since this unit is not in the same position relative to the other antennas in the array; (2) the transmission-line end-effects and coupling effects are different for the central unit from what they are for the two outer units.

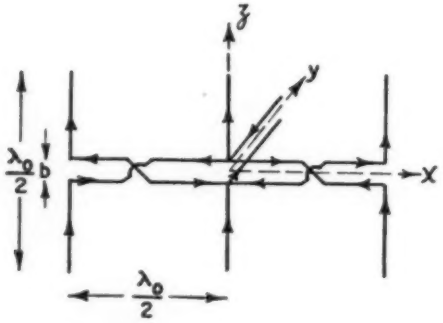


FIG. 3. Three-element broadside array.

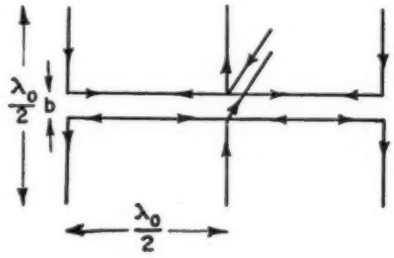


FIG. 4. Three-element bidirectional end-fire array.

Let the problem of transmission-line end-effects and coupling effects between each antenna and the feeding line be sidestepped in this paper by assuming that the distance b between the conductors of the line is sufficiently small to make these effects negligible. If this is done, formulas for the ratio of the current in antenna 1 or 3 (which have equal currents) to the current in the central unit 2 are obtained readily for $\beta_0 h = \pi/2$ since the simple formula (13) with $N=3$ is a good approximation. With negligible terminal-zone effects, the two sections of line of length $\lambda_0/2$ are one-to-one transformers so that the same voltage is applied to each of the three units. That is

$$\mathbf{V}_{30}=\mathbf{V}_{20}=\mathbf{V}_{10}. \tag{16}$$

Moreover, since antennas 1 and 3 are geometrically and electrically identical, it follows that

$$\mathbf{I}_{30}=\mathbf{I}_{10};\ \mathbf{Z}_{12}=\mathbf{Z}_{23}. \tag{17}$$

With (13) and (12) the following relations are obtained directly when $N=3$:

$$\mathbf{V}_{10}=\mathbf{V}_{20}=\mathbf{I}_{10}(\mathbf{Z}_{11}+\mathbf{Z}_{13})+\mathbf{I}_{20}\mathbf{Z}_{12}=2\mathbf{I}_{10}\mathbf{Z}_{12}+\mathbf{I}_{20}\mathbf{Z}_{22}. \tag{18}$$

The solution of (18) for the ratio of currents is

$$\mathbf{I}_{20}/\mathbf{I}_{10}=(\mathbf{Z}_{11}+\mathbf{Z}_{13}-2\mathbf{Z}_{12})/(\mathbf{Z}_{22}-\mathbf{Z}_{12}). \tag{19}$$

The expressions for the impedances are

$$\mathbf{Z}_{3in}=\mathbf{Z}_{1in}\equiv\frac{\mathbf{V}_{10}}{\mathbf{I}_{10}}=\mathbf{Z}_{11}+\mathbf{Z}_{13}+\mathbf{Z}_{12}\left[\frac{\mathbf{Z}_{11}+\mathbf{Z}_{13}-2\mathbf{Z}_{12}}{\mathbf{Z}_{22}-\mathbf{Z}_{12}}\right] \tag{20}$$

$$\mathbf{Z}_{2in}\equiv\frac{\mathbf{V}_{20}}{\mathbf{I}_{20}}=\mathbf{Z}_{22}+2\mathbf{Z}_{12}\left[\frac{\mathbf{Z}_{22}-\mathbf{Z}_{12}}{\mathbf{Z}_{11}+\mathbf{Z}_{12}-2\mathbf{Z}_{12}}\right]. \tag{21}$$

If all antennas are identical and have no series impedances,

$$\mathbf{Z}_{33}=\mathbf{Z}_{22}=\mathbf{Z}_{11}=\mathbf{Z}_{s1}. \tag{22}$$

In this case it is evident that $\mathbf{I}_{30}=\mathbf{I}_{10}$ must differ considerably from $\mathbf{I}_{20}$, and $\mathbf{Z}_{3in}=\mathbf{Z}_{1in}$ from $\mathbf{Z}_{2in}$ since $\mathbf{Z}_{13}$ and $\mathbf{Z}_{12}$ are far from the same. For example,

$$\Omega=10,\ b=\lambda_0/2,\ \mathbf{Z}_{s1}\doteq 73+j31;$$
$$\mathbf{Z}_{12}=-13-j26;\ \mathbf{Z}_{13}=3+j16. \tag{23}$$

It is possible to make the currents all equal when the driving voltages are the same by connecting an impedance in series with the central unit, antenna 2. Specifically, let

$$\mathbf{Z}_{33}=\mathbf{Z}_{11}=\mathbf{Z}_{s1};\ \mathbf{Z}_{22}=\mathbf{Z}_{s1}+\mathbf{Z}_2. \tag{24}$$

Upon setting the currents equal in (19) and using (24), the result is,

$$\mathbf{Z}_2=\mathbf{Z}_{13}-\mathbf{Z}_{12}. \tag{25}$$

It is seen that $\mathbf{Z}_2$ is quite a large impedance with a significant real part so that appreciable ohmic losses are involved which lower the efficiency of transmission. Since the electromagnetic field of a three-element array in which the current in the central unit is greater than that in the outer units may be more desirable than when the currents are all equal, it is seldom advantageous to use $\mathbf{Z}_2$ as specified in (25) in series with the central unit. An alternative procedure is to connect reactances in series with the *outer* units in order to make the phases of all three units the same. That is, in place of (24), let

$$\mathbf{Z}_{33}=\mathbf{Z}_{11}=\mathbf{Z}_{s1}+jX_1;\ \mathbf{Z}_{22}=\mathbf{Z}_{s1}. \tag{26}$$

The values in (16) now must be substituted in (20) and X_1 adjusted to make the expression real. That is,

$$\frac{\mathbf{I}_{20}}{\mathbf{I}_{10}}=\frac{(R_{s1}+R_{13}-2R_{12})+j(X_1+X_{s1}+X_{13}-2X_{12})}{(R_{s1}-R_{12})+j(X_{s1}-X_{12})} \tag{27}$$

must be real. This is true when

$$\frac{X_1+X_{s1}+X_{13}-2X_{12}}{R_{s1}+R_{13}-2R_{12}}=\frac{X_{s1}-X_{12}}{R_{s1}-R_{12}}, \tag{28a}$$

or when

$$X_1=\left(\frac{X_{s1}-X_{12}}{R_{s1}-R_{12}}\right)\times(R_{s1}+R_{13}-2R_{12})-X_{s1}-X_{13}+2X_{12}. \tag{28b}$$

Evidently

$$\frac{\mathbf{I}_{20}}{\mathbf{I}_{10}}=\frac{R_{s1}+R_{13}-2R_{12}}{R_{s1}-R_{12}}. \tag{29}$$

This ratio may be considerably greater than one.

Similar analyses may be carried out for arrays of any number of parallel units.

A Dimensional Analysis of Metal Cutting*

D. C. Drucker† and H. Ekstein‡
Armour Research Foundation, Chicago, Illinois
(Received June 30, 1949)

Dimensional analysis is used in a study of the behavior of metal when it is cut. Present theories are described and some additional significant variables are found. The analysis provides a basis for the interpretation of known experimental results, and should serve as a guide for future experimental and theoretical investigations.

INTRODUCTION

IF a problem is so complicated that a sufficiently comprehensive theory cannot be developed at the time, a very useful approach to a solution is to write down the independent variables of the problem and form their dimensionless combinations.[1] The significance of each independent dimensionless variable can then be determined from prior knowledge, reasoning, or experiment, and usually many of the dimensionless combinations can be ruled out as insignificant, that is, having no noticeable influence upon the result. Then, if the important independent combinations are the same, all dependent dimensionless variables will also be the same for both experimental and production runs. This means that practical predictions, such as of power requirement and chip continuity or surface finish, can be made from a limited set of experiments or field information without actually testing or ever having tested the proposed set-up. As the dimensionless combinations are not unique, a rearrangement is often worthwhile, for in this way the number of physically important variables may be materially reduced. This procedure has had widespread application in fluid mechanics and associated problems and is well standardized. Its use in mechanical problems is increasing as its value becomes more evident.[2]

Another, and more subtle, use of dimensional analysis is to check the physical completeness of a theory before embarking on the arduous task of grinding out mathematical formulas and numerical calculations. This, too, is illustrated in the following discussion and extension of present theories of metal cutting.

IDEALIZED METAL CUTTING

All theory must involve approximation to, or idealization of, actual behavior. Metal cutting offers a very difficult problem even for the two-dimensional case, Fig. 1, and the description proposed by Piispanen[3] and independently by Ernst and Merchant[4] was a tremendous forward step. The chip is assumed to come off continuously and there is no built-up edge. The fundamental action is a plastic shearing of each plane as it passes OA, the final appearance being similar to that which would be obtained with an inclined deck of very thin cards.

The resultant force R exerted by the tool on the chip per unit width of cut, Fig. 2, its horizontal or power component H, the shear angle φ, the chip thickness, and so forth, are dependent variables determined by the properties of the material being cut and by other independent variables. No one has yet devised a general theory which takes all into account. The major question is which variables are most important and how do they enter. In the simplest theory it is assumed that the independent variables are only the rake angle α, the depth of cut t, the cutting speed V, a constant friction angle between tool and chip β, and a fixed yield stress in shear τ_y. Once this initial step has been taken, R, φ, H, and so forth can be found; Ernst and Merchant have written down the results. However, a dimensional analysis will show that despite its great value there are practical physical limitations for any theory, elaborate or simple, of the form

$$\varphi \text{ or } H \text{ a function of } \alpha,\ t,\ V,\ \beta,\ \tau_y.$$

φ is dimensionless as are α, and β, but H is a force per

---

* Based on a research program sponsored by Shell Oil Company, Inc., and performed at Armour Research Foundation, Chicago, Illinois.

† Formerly on the staff of Armour Research Foundation, Chicago, Illinois. Now Associate Professor of Engineering, Brown University, Providence, Rhode Island.

‡ Senior Physicist, Physics Research Department, Armour Research Foundation, Chicago, Illinois.

[1] The reviewer of this paper brought to the authors' attention the excellent article by F. F. P. Bisacre and G. H. Bisacre, "Life of carbide-tipped turning tools," Proc. Inst. Mech. Eng. War Emergency Issue No. 35, **157**, 452–469 (1947). A dimensional analysis is included which in certain respects is more complete than ours. However the approach and the objective are quite different and a few variables important here are not included. The reviewer also pointed out that Dr. B. T. Chao has submitted a paper to the Institution of Mechanical Engineers in which, following Bisacre and Bisacre, considerable use is made of $K/Vc\rho t$.

[2] For a particular example in which there are very many independent variables but only one is found to be of real importance in the practical range, see D. C. Drucker and H. Tachau, "A new design criterion for wire rope," J. App. Mech. **XII**, No. 1, A-33–A-38 (1945).

[3] V. Piispanen, "Lastunmodostumisen teoriaa (Theory of chip formation)," Teknillinen Aikakauslehti **XXVII**, No. 9, 315–322 (1937).

[4] H. Ernst and M. E. Merchant, "Chip formation, friction, and finish," The Cincinnati Milling Machine Company (1940); (also "Surface treatment of metals," Symposium, Trans. A.S.M.E., 299–378 (1941); M. E. Merchant, "Mechanics of the metal cutting process"; I. "Orthogonal cutting and a type 2 chip," J. App. Phys., **XVI**, No. 5, 267–275 (1945); II. "Plasticity conditions in orthogonal cutting," J. App. Phys., **XVI**, No. 6, 318–324 (1945).

unit length, t is a length, V is a length over time, and τ_y is a force divided by length squared. The first step is the formation of the independent dimensionless combinations of the independent variables. It is easily seen that only α and β remain. V alone contains time and τ_y alone contains units of force so that a dimensionless combination cannot be formed with either, and both must be eliminated. Then t remains as the only independent variable involving length, and it must also be dropped. Therefore, any dimensionless quantity containing a dependent variable must turn out to be a function of α and β only, for example,

$$\varphi=\varphi(\alpha, \beta)$$

and

$$H/t\tau_y=H/t\tau_y(\alpha, \beta).$$

This is illustrated by the formulas given by Ernst and Merchant for these conditions, for example,

$$\varphi=45°-(\beta-\alpha)/2.$$

More elaborate theories are possible and other results can be obtained, but if they are based upon the same five independent quantities only, no formula developed can contain either the speed or the depth of cut as independent variables. As both are known to be important in several ways, such theories, although very useful, are definitely incomplete. Also, the predictions for different materials appear too similar. The calculated shear angle φ will be the same if the friction and rake angles are the same, no matter what the properties of the material.

The most general theory now in use includes the material to the extent of considering the variation of the yield stress in shear with the normal stress σ on the shear plane. The change is generally negligible at ordinary stress magnitudes, but the stresses encountered in metal cutting are extremely high, exceeding 200,000 p.s.i. for mild steel. If the relation is linear, of the form $\tau_y=\tau_0+k\sigma$, the previous dimensional analysis must be modified by replacing τ_y by the new independent variables τ_0 and k. As k is dimensionless, the result can be written immediately as

$$\phi=\phi(\alpha, \beta k)$$
$$H/t\tau_0=H/t\tau_0(\alpha, \beta, k)$$

which is shown by Merchant's modified expressions,[4] for example

$$\phi=C/2-(\beta-\alpha)/2,$$

where C is arc *ctnk*. There is still no influence of depth or speed of cut.

NEW CONCEPTS AND VARIABLES

The necessity for new concepts and the accompanying variables is thus made clear by the dimensional reasoning. It is known that as the speed is decreased or as the depth is increased, the chip will gradually change from the continuous type to the segmental. Present theories therefore cannot indicate under what conditions they break down, that is, when the picture they postulate is no longer valid. Furthermore, the shear angle and the specific cutting force generally do vary with V and t, which was shown to be incompatible with any theory using only the variables mentioned above. Therefore, there is a strong suggestion that a variable T_0 with the dimension of time be introduced. Such an independent variable with physical meaning would include the cutting speed by adding the dimensionless group VT_0/t. T_0 could be a time required for the material to fracture. It would not be valid to introduce such a variable unless T_0 would be of the order of magnitude of the essential time interval in the cutting operation. However, a simple calculation shows that the time in which the shearing action takes place in actual cutting is extremely short and furthermore the time in the ideal case, Fig. 1, is zero. Thus, dimensional analysis can lead to new and useful concepts; in metal cutting the transition from the continuous to the discontinuous chip apparently is governed by the time available for fracture. The transition value of VT_0/t will, of course, depend on the shear strain, which in turn depends upon α and β only. If the additional variables which will be considered later are of minor significance, then the dimensional analysis shows that a given percentage increase in t requires the same percentage increase in V to obtain the same degree of chip continuity and, therefore, approximately the same surface finish.

Another important omission in the idealized picture of cutting is the lack of prediction of any effect of depth of cut on the specific cutting force. It is always observed in practice that when t is small, the value of $H/t\tau_0$ rises very sharply as t decreases. This fact leads to the consideration of lengths which might be significant in cutting. The nose radius of the tool r_N, which was assumed as zero, could account for this phenomenon qualitatively and would appear in the variable r_N/t.

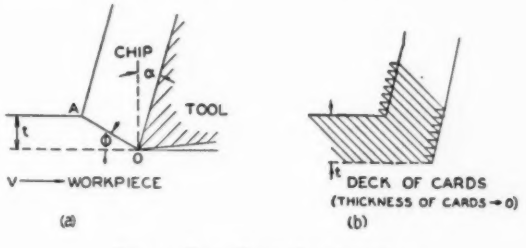


FIG. 1. Two-dimensional cutting.

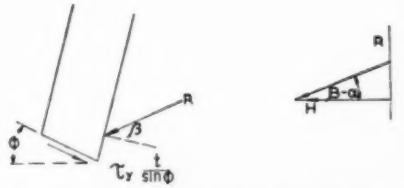


FIG. 2. Forces on chip of unit width (perpendicular to paper).

TABLE I

Variable	Units	Description
c	$FLM^{-1}\theta^{-1}/J$	specific heat, for example, B.t.u./lb.mass/deg.F
d	L	spacing of disturbances in crystal structure
G	FL^{-2}	shear modulus of elasticity
K	$FT^{-1}\theta^{-1}/J$	thermal conductivity, for example, B.t.u./sq.ft./deg.F/hr. for a unit thickness.
k	none	slope of curve of yield point in shear *vs.* normal stress on the shear plane
r_N	L	nose radius of tool
T_0	T	time required for fracture
t	L	depth of cut
V	LT^{-1}	cutting speed—relative speed of tool and workpiece
α	none	rake angle
γ_L	none	ductility—a limiting value of shearing strain
μ	none	a coefficient of rubbing friction—not necessarily simply tan β described previously
ψ_t	θ	a transition temperature of the metal, for example, melting point
ψ_w	θ	workpiece temperature
ρ	ML^{-3}	mass density
τ_0	FL^{-2}	yield stress in shear

However, it is found that the major part of the depth effect is independent of this ratio; a sharp tool taking a shallow cut is not equivalent to a more rounded tool at larger depths but instead will give a much higher specific cutting force. The effect is, therefore, apparently not one that would be observed in a perfectly homogeneous material. A grain size or, possibly better, a mean distance between imperfections of the metal, seems to provide a reasonable choice.

The preceding additional variables might be termed borderline because they apply or become important when the ideal picture starts to break down visibly. While they are very important, it is also true that the experimental results often do not agree well in detail with the present limited theory even when the chip is definitely continuous and the depth of cut is large enough to avoid the equivalent of a size effect. This is not surprising because no mention has been made so far of temperature changes, which are always large locally and usually have an appreciable influence throughout. The tangential force between tool and chip is certainly markedly dependent upon temperature; the yield point in shear will also be affected although the enormously high strain rate may nullify much of the temperature change; ductility will be radically altered; and so on. Again dimensional analysis can point the way.

THE GENERAL CASE

For the analysis of the general case, it is desirable to list all of the independent variables that are to be considered, with their units of force F or mass M, length L, time T, and temperature θ. Both F and M will be used for convenience, $F \equiv MLT^{-2}$. The mechanical equivalent of heat, J, is included to emphasize that a conversion of heat to mechanical units will be necessary at times (see Table I).

If this were a complete list of the independent variables, a functional form of the solution could be obtained easily. There are sixteen quantities and four fundamental units so that twelve independent dimensionless groups exist. If, for example, the shear angle φ, the horizontal force per unit width of cut H, or the temperature rise at the shear plane ψ, is being investigated, then φ, $H/t\tau_0$, or $\psi/(\psi_t-\psi_w)$ can be expressed as a function of $[\alpha, \mu, \gamma_L, k, VT_0/t, r_N/t, d/t, \tau_0/G, \rho V^2/\tau_0, K/Vc\rho t, \tau_0/Jc\rho(\psi_t-\psi_w), \psi_t/\psi_w]$.

The groupings are arbitrary within a very wide choice. Some of the dimensionless variables have been discussed already, the others were chosen because physical meaning could be attached to them easily. The variables α and μ which appear in the idealized theory are definitely important, and k will always have influence. Borderline variables VT_0/t, d/t, and γ_L become important as the continuity of the chip decreases. The ratio r_N/t can usually be ignored for a properly sharpened tool.

More analysis is required for the temperature variables. The shear strain (a function of the independent variables) multiplied by $\tau_0/Jc\rho$ would be the temperature rise at the shear plane if k were zero and if the heat generated were convected but not otherwise conducted away by the metal. $K/Vc\rho t$ is essentially the ratio of the heat conducted from the shear plane to the heat convected by the moving metal. If the temperature rise at the shear plane should be small compared with all significant temperature rises, $\psi_t-\psi_w$, for the metal being cut, then the three-temperature variables can be neglected.

The two remaining groups would seem to be borderline variables, but controlled experimentation is necessary before a definite statement can be made. τ_0/G is the maximum elastic shear strain (the resilience could be used in a somewhat more complicated group). $\rho V^2/\tau_0$ is essentially the ratio of inertia stress to the yield stress in shear and is always extremely small.

So far, the positive aspects of dimensional analysis have been emphasized. Unfortunately, it, too, has its drawbacks. The list of variables may not be sufficiently complete, nor, what is far worse, truly independent. The matter of completeness is rarely obvious, significant omissions may show up when an experimental program cannot be correlated with the use of variables already written down. However, this same trouble must be cleared up before an analytical approach can be successful.

Some of the difficulties can be seen by a re-examination of the list. Quantities like specific heat, thermal conductivity, ductility, coefficient of friction, and yield stress vary not only from one material to the other but, unpleasantly, also in a given material as the cutting conditions vary. The symbols must therefore be interpreted as the (fixed) values for a given material under a specified set of conditions. If these initial values

and the remaining variables completely determine the values for all other sets of conditions, then the list of variables is complete. This will require at the very least a number of coefficients similar to k, for example, a set of constants for the non-linear variation of specific heat and of thermal conductivity with temperature. If independent quantities of this type exist, no matter how many there may be, there is no essential additional complication as far as an experimental program is concerned although theoretical solutions may be practically ruled out. The metal being cut enters in so many ways that only in the simplest fundamental work will it be possible to substitute other materials, as can very often be done in fluid mechanics problems. Therefore, in most cases, the large number of required constants which should be added to the list of variables will automatically be the same in model and prototype, experiment and production, or two production runs. In fact, some of the dimensionless combinations already considered are similarly fixed. Therefore, at appreciable depths of cut, probably only

$$(\alpha, VT_0/t, K/Vc\rho t)$$

are of major significance *for a given material and temperature of workpiece.*

It is interesting that almost as in the Reynolds and Froude numbers for fluids, V and t appear in ratio in one-dimensionless variable and as a product in the other. Therefore, complete similarity between two experiments cannot be obtained using one material and is rarely practicable. This discouraging result is almost always found when the list of variables is reasonably extensive. However, it may be true, as for fluids, that the range of importance of each dimensionless group containing V, t is markedly separate, or that the phenomena they control are quite distinct. There is a strong probability that this is true for metal cutting. VT_0/t seems to control the degree of continuity of the chip while $K/Vc\rho t$ should be of major importance in the definitely continuous range. Only careful experimental investigations of many different metals will provide the answer.

In the previous discussion, the tool itself has hardly been considered. Here, again, it is likely that in only a few cases can information be obtained with one tool material and applied to another in a quantitative manner. Also, the use of a cutting fluid or coolant would add many variables. All of these difficulties merely emphasize the impossibility at present of developing a mathematical theory which could quantitatively cover even a small section of the complete problem, and demonstrate the necessity of carefully conceived experimental investigations based on dimensional analysis.

CONCLUSION

The best procedure to obtain a clear picture of metal cutting would seem to be an experimental analysis of dry two-dimensional cutting on the basis of the dimensionless variables already discussed. Despite the very large volume of data on the cutting of metals there is only a little information that is directly useful for this particular fundamental purpose.

The two main variables, V and t, appearing in VT_0/t and $K/Vc\rho t$, are of prime significance because they can be varied in a known and simple manner, but most of the others are of equal physical importance and may merit careful study. Those involving workpiece temperature directly can also be varied without too much difficulty and offer interesting possibilities.

It is quite likely that additional variables will be found necessary to explain all results and that a rearrangement of some of the suggested dimensionless groups will be advantageous.

ACKNOWLEDGMENT

The authors wish to thank the Shell Oil Company for permission to publish this paper.

The Current Build-Up in a Planar Diode

EDWARD H. GAMBLE*
Polytechnic Institute of Brooklyn, Brooklyn, New York
(Received June 29, 1949)

A theoretical study has been made of the initial electron flow when a signal voltage $V(t)$ is applied across an idealized planar diode. The analysis is based upon three fundamental postulates: Poisson's equation, Newton's equation, and the continuity equation. This quasi-stationary study includes both the growth and motion of the space-charge cloud of electrons across the interelectrode space and the transient build-up of the current in the external circuit. The analytical forms and numerical plots of the current build-up under space-charge-limited and temperature-limited operation are compared for various "large" and "small" signal voltages.

STATEMENT OF THE PROBLEM

A STUDY of the initial electron flow when a signal voltage $V(t)$ is applied across a planar diode has been undertaken. This problem may be attacked from two possible directions. One can make a study of the way in which the space-charge cloud of electrons forms and determine the motion of this charge under the action of the electric field due to the applied signal and the counter effect of the charge upon this field. The other approach is that of studying the transient build-up of the current in the external circuit.

This latter approach is the one which will be adopted to secure the immediate results of interest to the electronics engineer. By analogy to hydrodynamics, one may specify the quantities describing the moving electron either in the Lagrange variables (t,τ) or the Euler variables $[t,x]$. t is the clock time measured from the instant when the voltage is applied to the diode, and τ denotes the instant when a particular group of electrons is emitted from the cathode surface. τ is referred to as the "emission" time.

The diode configuration is shown in Fig. 1.

DETERMINATION OF THE DENSITY OF THE SPACE-CHARGE CLOUD

The symbol Φ denotes the scalar potential, E the electric field density, and ρ the space-charge density.

If one makes the assumption that any signal voltages which one may apply across the electrodes of the vacuum diode do not change significantly in the time required for an electromagnetic wave to propagate the interelectrode space, one may treat the entire problem as quasi-stationary. The three basic electronic equations reduce to Poisson's equation, Euler or Newton's equation of motion, and the continuity equation. The formulation of the important physical quantities then becomes

$$E[t,x]=-\frac{V(t)}{h}+\frac{1}{h\epsilon_0}\int_0^h(h-\xi)\rho[t,\xi]d\xi-\frac{1}{\epsilon_0}\int_0^x\rho[t,\xi]d\xi \quad (1)$$

$$\Phi[t,x]=\frac{x}{h}V(t)-\frac{x}{h\epsilon_0}\int_0^h(h-\xi)\rho[t,\xi]d\xi+\frac{1}{\epsilon_0}\int_0^x(x-\xi)\rho[t,\xi]d\xi \quad (2)$$

$$I(t)=\frac{\epsilon_0}{h}\frac{d}{dt}V(t)+\frac{1}{h}\int_0^h\left\{-\rho[t,x]\frac{\partial x}{\partial t}\right\}dx \quad \text{and}$$

$$=\frac{\epsilon_0}{h}\dot{V}(t)+I_{\text{ind}}(t), \quad (3)$$

where the subscript *ind* signifies that the quantity is related to the current induced in the external circuit due to the motion of the electrons.

$$x(t,\tau)=u(\tau,\tau)[t-\tau]+\frac{e}{mh}\int_\tau^t(t-\eta)V(\eta)d\eta-\frac{e}{mh\epsilon_0}\int_\tau^t(t-\eta)\left\{\int_0^h(h-\xi)\rho[\eta,\xi]d\xi\right\}d\eta+\frac{e}{m\epsilon_0}\int_\tau^t(t-\eta)\left\{\int_0^x\rho[\eta,\xi]d\xi\right\}d\eta, \quad (4)$$

where $u(\tau,\tau)$ denotes the initial velocity of the τ-group of electrons.

If one restricts the problem to the initial electron flow and hence the period of time to that of the transit time of the first electron,

$$0<t<t_1, \quad \tau=0, \quad (5)$$

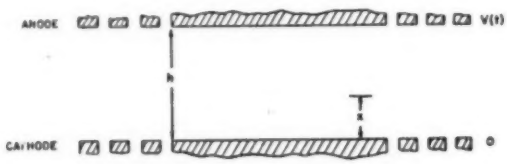


FIG. 1. Planar diode configuration.

* Now with Curtiss-Wright Corporation, Columbus, Ohio.

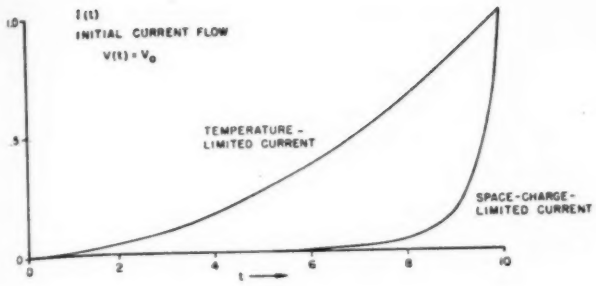


FIG. 2.

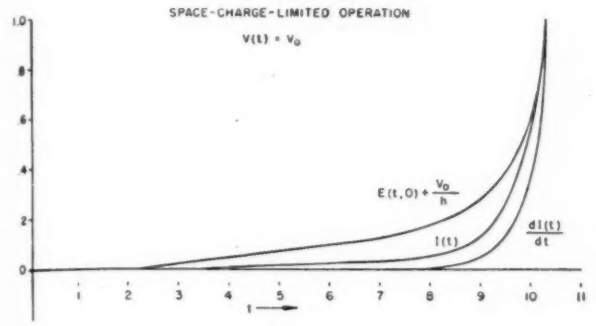


FIG. 3.

Eq. (4) becomes

$$x(t,0)=u(0,0)t+\frac{e}{mh}\int_0^t(t-\eta)V(\eta)d\eta$$
$$-\frac{e}{mh\epsilon_0}\int_0^t(t-\eta)\left\{\int_0^h(h-\xi)\rho[\eta,\xi]d\xi\right\}d\eta$$
$$+\frac{e}{m\epsilon_0}\int_0^t(t-\eta)\left\{\int_0^x\rho[\eta,\xi]d\xi\right\}d\eta. \quad (6)$$

Equation (6) may be considered as an integral equation for $\rho[t,x]$ when one specifies the position x and the time t. Knowing the form of $\rho[t,x]$, this quantity may be substituted into Eqs. (1)–(3) to determine the other desired quantities.

DETERMINATION OF THE EXTERNAL CURRENT

If one chooses the total current $I(t)$ as the unknown quantity to be found, the important equations are

$$E(t,\tau)=E_c(\tau,\tau)-\frac{V(t)}{h}-\frac{1}{\epsilon_0}\{q_{\rm ind}(t)-q_{\rm ind}(\tau)\}$$
$$=E_c(\tau,\tau)-\frac{V(t)}{h}-\frac{1}{\epsilon_0}\int_\tau^t I_{\rm ind}(\eta)d\eta, \quad (7)$$

where $E_c(\tau,\tau)$ is the field at the cathode at the instant τ, and

$$q_{\rm ind}(t)=\int_0^t I_{\rm ind}(\eta)d\eta, \quad (8)$$

$$\Phi(t,\tau)=-\int_\tau^t E(t,\eta)\frac{\partial x(t,\eta)}{\partial\eta}d\eta$$
$$=-E_c(\tau,\tau)x(t,\tau)+\frac{x(t,\tau)}{\epsilon_0}\int_\tau^t I(\eta)d\eta$$
$$+\int_\tau^t x(t,\eta)\frac{\partial E(\eta)}{\partial\eta}d\eta, \quad (9)$$

$$x(t,\tau)=u(\tau,\tau)[t-\tau]-\frac{e}{2m}E_c(\tau,\tau)[t-\tau]^2$$
$$+\frac{e}{m\epsilon_0}\int_\tau^t\frac{(t-\eta)^2}{2}I(\eta)d\eta$$
$$=u(\tau,\tau)[t-\tau]-\frac{e}{2m}E_c(\tau,\tau)[t-\tau]^2$$
$$+\frac{e}{mh}\int_\tau^t(t-\eta)V(\eta)d\eta$$
$$+\frac{e}{m\epsilon_0}\int_\tau^t\frac{(t-\eta)^2}{2}I_{\rm ind}(\eta)d\eta. \quad (10)$$

SPACE-CHARGE-LIMITED CURRENT BUILD-UP

If one applies the boundary conditions

$$E_c(\tau,\tau)=E[t,0]=0, \quad u(\tau,\tau)=0, \quad (11)$$

the relations given above reduce to those found previously by Wang[1] for the space-charge-limited planar diode. For conditions (5), the basic integral equation becomes

$$V(t)=\frac{h}{\epsilon_0}q(t)-\frac{e}{m\epsilon_0^2}\int_0^t\frac{(t-\eta)^2}{2}q(\eta)I(\eta)d\eta. \quad (12)$$

Differentiating Eq. (12) twice with respect to t, one finds that

$$\frac{\partial^2V}{\partial t^2}=\frac{h}{\epsilon_0}\ddot{q}(t)+\frac{e}{2mh}q^2(t). \quad (13)$$

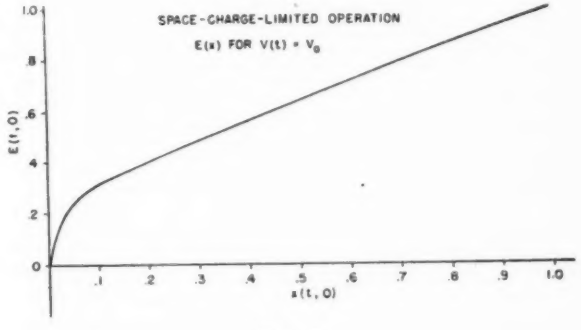


FIG. 4.

[1] C. C. Wang, "Large signal high frequency electronics of thermionic vacuum tubes," Proc. I. R. E. (April, 1941).

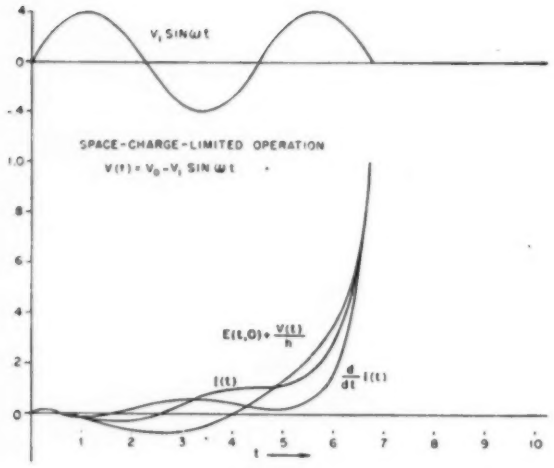


FIG. 5.

The solution of Eq. (13) for $(d/dt)q(t)=I(t)$ will give the current build-up for the space-charge-limited diode during this initial period before the electrons strike the anode.

TEMPERATURE-LIMITED CURRENT BUILD-UP

The conditions for temperature-limited operation may be defined as

$$I_0=\text{saturation current}=\left\{-\rho[t,x]\frac{\partial x}{\partial t}\right\} \tag{14}$$

$$E[t,x]=-V(t)/h=E(t) \tag{15}$$

$$\Phi[t,x]=[x(t,\tau)/h]V(t). \tag{16}$$

For the period before any electrons strike the anode,

$$x(t,\tau)=u(\tau,\tau)[t-\tau]+\frac{e}{mh}\int_\tau^t(t-\eta)V(\eta)d\eta \tag{17}$$

$$\Phi(t,\tau)=\frac{u(\tau,\tau)}{h}[t-\tau]V(t)+\frac{V(t)}{mh^2}\int_\tau^t(t-\eta)V(\eta)d\eta \tag{18}$$

$$I(t)=\frac{\epsilon_0}{h}\dot{V}(t)+\frac{I_0}{h}u(\tau,\tau)[t-\tau]+\frac{I_0e}{mh^2}\int_\tau^t(t-\eta)V(\eta)d\eta. \tag{19}$$

SOME SPECIFIC CASES

We now have all of the basic formulas necessary to determine the initial current build-up. It will be advantageous to compare the space-charge-limited and the temperature-limited cases. These are the solutions of Eqs. (13) and (19), respectively, for conditions (5).

Let $V(t)=V_0$, so that the build-up for a constant applied signal with its application to pulses may be determined. Figure 2 shows an enlarged plot of the solutions for this case. This plot shows the final current rising to the same value for both classes of operation. This will, in general, not be true. The functional form of the current as a function of time is of importance.

The temperature-limited current according to Eq. (19) is

$$I_{\text{ind}}(t)=(I_0e/mh^2)V_0(t^2/2) \tag{20}$$

for negligible initial velocity $u(\tau,\tau)$.

The position of the leading edge of the electron cloud is given by Eq. (17).

$$x(t,0)=u(\tau,\tau)\cdot t+(V_0/2mh)t^2. \tag{21}$$

The current and space charge always build up as the same power of time t for the temperature-limited operation.[2]

The numerical solution of Eq. (13) indicates that the space-charge-limited current rises much more rapidly than a square of the clock time. An accurate empirical formula for this current is given by Eq. (22)

$$I_{\text{ind}}(t)=I_{\text{const}}\left\{30.121\left(\frac{t}{t_1}\right)^{10}-51.009\left(\frac{t}{t_1}\right)^{9}+21.889\left(\frac{t}{t_1}\right)^{8}\right\}. \tag{22}$$

The complete analytical solution of Eq. (13) for this case is an elliptic integral

$$I_{\text{ind}}(t)=\dot{q}_{\text{ind}}(t)=\int_0^t\left[\int_0^\eta I_{\text{ind}}(\xi)d\xi\right]^2 d\eta. \tag{23}$$

For this case, Eq. (7) reduces to

$$E(t,0)+\frac{V_0}{h}=-\frac{q_{\text{ind}}(t)}{\epsilon_0}=-\frac{1}{\epsilon_0}\int_0^t I(\eta)d\eta=-\frac{1}{\epsilon_0}I_{\text{const}}\left\{39.405\left(\frac{t}{t_1}\right)^{11}-73.406\left(\frac{t}{t_1}\right)^{10}+35.001\left(\frac{t}{t_1}\right)^{9}\right\}. \tag{24}$$

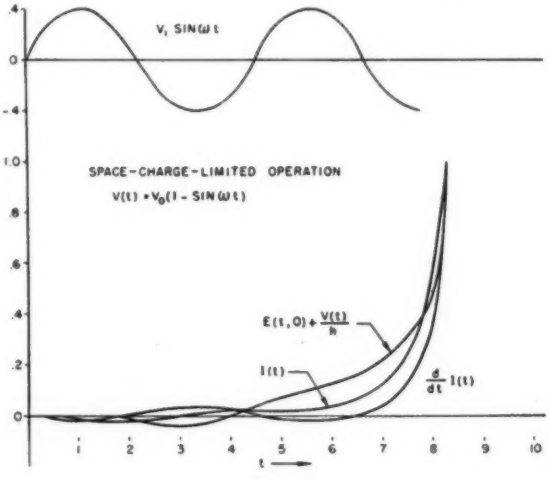


FIG. 6.

[2] B. Salzberg, "Large signal high frequency theory of the vacuum tube," dissertation for D.E.E. at the Polytechnic Institute of Brooklyn, Brooklyn, New York (1941).

Hence, a plot of $q(t)$, as shown in Fig. 3, is also a plot of the electric field as a function of time. For the determination of the leading edge of the electron cloud, one must employ Eq. (10)

$$x(t,0)=\frac{e}{2m\epsilon_0}30.121I_{\text{const}}\int_{\tau=0}^{t}(t-\eta)^2F(\eta)d\eta, \quad (25)$$

where

$$F(\eta)=\left[\frac{(t-\eta)^{10}}{t_1^{10}}-1.6935\frac{(t-\eta)^9}{t_1^9}+0.7267\frac{(t-\eta)^8}{t_1^8}\right]$$

$$x(t,0)=\frac{e}{m\epsilon_0}15.06I_{\text{const}}\left[\left(\frac{t}{t_1}\right)^{10}-1.6935\left(\frac{t}{t_1}\right)^9-0.7267\left(\frac{t}{t_1}\right)^8\right]t^2. \quad (26)$$

A plot can be made of $E(x)$ by a comparison of $E(t,0)$ given by the uppermost plot of Fig. 3 and the data computed from Eq. (26). This is shown in Fig. 4.

For the applied signal

$$V(t)=V_0-V_1\sin\omega t \quad (27)$$

for

$$V_1>V_0, \quad \omega>1/t_1, \quad (28)$$

the functional forms of $I(t)$ and $(d/dt)I(t)$ are shown in Fig. 5. The function

$$[E(t,0)+(V(t)/h)]$$

is also plotted in Fig. 5. The electric field $E(t,0)$ for the applied voltage given by Eq. (27) is

$$E(t,0)=-\frac{V_0}{h}+\frac{V_1\sin\omega t}{h}-\frac{1}{\epsilon_0}\int_{\tau=0}^{t}I_{\text{ind}}(\eta)d\eta. \quad (29)$$

The actual value of the ratio V_0/V_1 is not determinable from the numerical solution of Eq. (13) since only the second time derivative of the applied signal

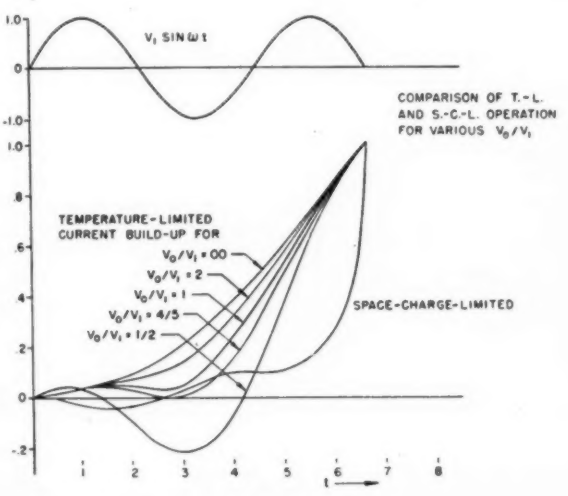


FIG. 7.

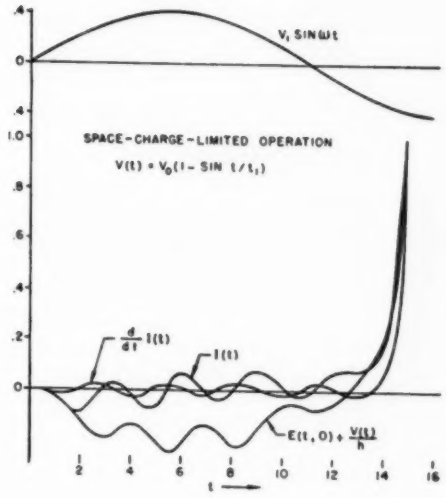


FIG. 8.

enters into the equation. However, experience has shown that if this ratio becomes too much smaller than *one*, the field will be negative and the electrons will never reach the anode. The period of the applied signal is clearly shown on the graph in order that one may visualize the time sequence. For the particular case where $V_0/V_1=1$, Fig. 6 shows the form of the current and the electric field as functions of time. A comparison of Figs. 5 and 6 for $I(t)$ indicates that the period of time during which the current is constant has been lengthened and the build-up time extended when the full modulation effect takes place for the amplitude ratio *one*.

It was considered advisable to make a direct comparison of the current build-up for space-charge-limited operation and temperature-limited operation for the applied signal given by Eq. (27) and conditions (28). Figure 7 shows this comparison. Several values of the ratio V_0/V_1 were chosen and plots made. The formula for the temperature-limited current is found by the substitution of Eq. (27) into Eq. (19).

$$I(t)=V_1\left\{-\frac{\epsilon_0}{h}\omega\cos\omega t+\frac{I_0e}{mh^2}\left[\frac{V_0}{V_1}\frac{t^2}{2}-\frac{t}{\omega^2}+\frac{\sin\omega t}{\omega^2}\right]\right\}. \quad (30)$$

The uppermost plot of Fig. 7 shows directly the comparison of the space-charge-limited current and the temperature-limited current for the case $V_0/V_1=\frac{1}{2}$. Another interesting plot is given in Fig. 7 which shows the variation of the build-up current with changes of the ratio V_0/V_1. This covers most of the range of large signals.

When the frequency of the time-varying component of the signal voltage is nearly equal to the reciprocal of the transit time, one finds that the build-up time is lengthened since the electric field is negative for a longer period of time. One point of interest is the large difference in the period of oscillation of the electric field and the current and the period of the signal com-

ponent. This is clearly shown in Fig. 8. The basic form of the signal voltage is

$$V(t)=V_0(1-\sin t/t_1). \tag{31}$$

CONCLUSIONS

The initial current build-up of thermionic vacuum tubes under space-charge-limited operation is a function of higher powers of time than under temperature-limited operation. The build-up may be considerably delayed or shortened when sinusoidally varying signals whose amplitude is comparable to the constant signal voltage are applied. The results shown in this paper should be of particular interest to engineers studying microsecond pulsing of thermionic vacuum tubes.

ACKNOWLEDGMENT

The author wishes to express his gratitude to the ONR who sponsored a special fellowship which enabled the author to carry on the research work given in this paper. A vote of thanks is due to the Reeves Instrument Company which made their analog computer available for numerical solutions. The author shall always be indebted to Dr. Ernst Weber of the Microwave Research Institute whose kindly advice and supervision served as a driving force and inspiration.

Thermocouples of the Refractory Metals

F. Haydn Morgan and W. E. Danforth

Bartol Research Foundation of The Franklin Institute, Swarthmore, Pennsylvania

(Received July 13, 1949)

The thermoelectric power of several refractory thermocouples was investigated, W-Ta for temperature measurements up to 3000°C, W-Mo up to 2600°C, Ta-Mo up to 2600°C and W-W/Mo (alloy 50 percent Mo/50 percent W) up to 2900°C.

Basic calibration was carried out using several melting points and the results were verified and interpolated by means of an optical pyrometer.

The couples W-Ta and W-Mo have been applied in this laboratory to the temperature measurement requirements in high temperature vacuum furnaces and determinations of spectral emissivity.

IN order to satisfy the need for accurate measurements of high temperature in electronic research, the thermocouples of the refractory metals have been investigated. Existing literature did not exceed 2000°C, with the exception of Morugina,[1] who studied W-Ta; and Pirani and Wagenheim,[2] who reported on W-W/Mo (75 percent W/25 percent Mo).

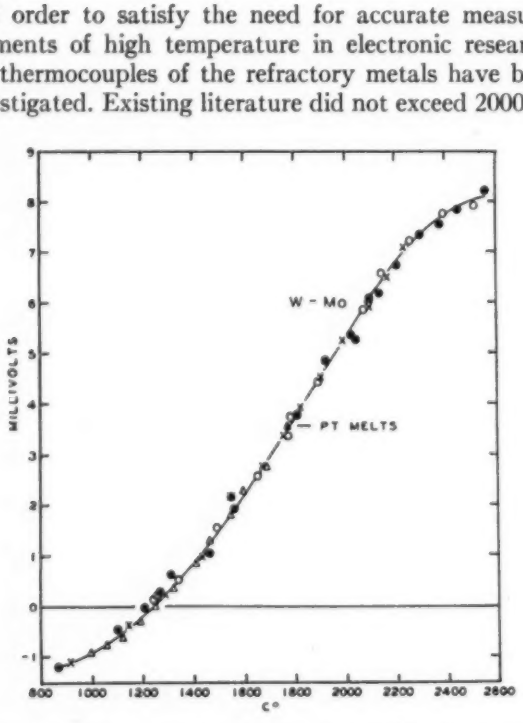


FIG. 1. Composite calibration data for W-Mo thermocouple. The different types of points represent different runs with optical pyrometer. The solid triangle represents melting point data taken with platinum in a vacuum furnace.

For the optical pyrometer comparisons the thermocouple wires 0.005 in. or 0.002 in. diameter were spot-welded to a 0.030-in. tungsten wire, which was heated by passage of current. This was placed in a bell jar and evacuated to 10^{-5} mm Hg. A Leeds and Northrup Type K potentiometer was used to measure the e.m.f. and a Leeds and Northrup Optical Pyrometer to measure the brightness temperature. Using 0.43 as the spectral emissivity[3] the brightness temperature was converted to true temperature. The cold junction was held at room temperature.

The melting point calibrations were obtained by spot-welding the thermocouple to the midpoint of a 2-cm length of 0.030-in. wire of the standard metal. The following metals were used:[4] Gold 1063.0°C ±0.0°, Nickel 1455°C±1°, Platinum 1773.5°C±1°, Molybdenum 2625°C±50°, and Tantalum 3000°C ±100°. An optical pyrometer would detect no temperature differences in the vicinity of the couple. The po-

[1] S. Morugina, Zeits. f. Tech. Physik **7**, 486 (1926).

[2] M. Pirani and G. Wagenheim, Zeits. f. Tech. Physik **6**, 358 (1925).

[3] W. F. Roeser and H. T. Wensel, *Handbook of Chemistry and Physics* (1947), thirtieth edition, p. 2246.

[4] National Bureau of Standards Miscellaneous Publication M183.

TABLE I. E.m.f. (millivolts).

T(°C)	Ta-Mo	W-Ta	W-Mo	W-W/Mo
800	12.20	10.65	−1.32	−0.50
900	13.60	12.20	−1.15	−0.30
1000	14.80	13.72	−0.80	0.00
1063		14.70		
1100	15.95	15.25	−0.60	0.30
1200	16.90	16.65	−0.20	0.70
1250	17.32	17.32	0.00	
1300	17.70	17.95	0.30	1.15
1400	18.35	19.05	0.85	1.65
1455		19.65		
1500	18.80	20.05	1.50	2.25
1600	19.25	20.90	2.18	2.85
1700	19.45	21.65	2.95	3.45
1773		22.12	3.54	
1800	19.45	22.25	3.73	4.00
1900	19.25	22.70	4.50	4.60
2000	18.85	22.90	5.30	5.20
2100	18.35		6.05	5.80
2200	17.70		6.80	6.40
2300	16.90		7.35	6.90
2400	15.90		7.75	7.35
2500	14.90		8.00	7.70
2600				7.95
2625		20.90		
2700				8.08
2800				8.12
2900				8.18
3000		18.50		

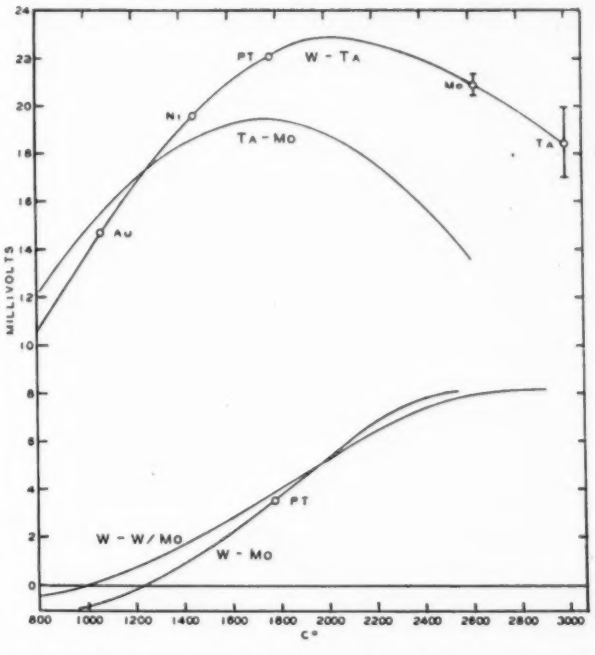


FIG. 2. Calibration curves of four refractory thermocouples. Melting point data are indicated on the W-Ta and W-Mo curves.

tentiometer was kept in balance as the temperature was slowly raised. When the wire melted, permitting the thermocouple to cool, the potentiometer would suddenly go out of balance. The last point of balance was taken as the melting point.

Considerable time was spent verifying the reproducibility of the present results. Wire stock from different spools and of different sizes was tested. Figure 1 shows composite calibration results in the case of the W-Mo couple. Similar sets of data were taken with the other couples and the resulting average curves are plotted in Fig. 2 and presented in Table I.

W-Ta (contrary to Morugina's findings of a straight line) gives a curve which rises from 14 millivolts at 1000°C to a maximum of 22.9 millivolts at 2000°C and then decreases as the temperature is increased. The slope of the curve between 1300°C and 1800°C is 8 microvolts per degree. The couple was found to be reproducible up to 2000°C, for which the graph shows pyrometer data. Above this temperature the pyrometer data showed excessive spread and, for some reason that was not determined, did not agree with the melting points of Mo and Ta which have been plotted.

The W-Mo curve has negative values below 1250°C, where it passes through zero and then rises to 8 millivolts at 2500°C. The average slope of the curve between these temperatures is 6.5 microvolts per degree. The scatter of the data in this case was relatively small, and agrees with literature published by B. Osann and Schroeder,[5] which gave the W-Mo curve up to 1850°C.

The W-W/Mo (50 percent W–50 percent M) curve is of very similar form to the W-Mo curve. It passes through zero at 1000°C and rises to 8 millivolts at 2600°C where it tends to saturate. The slope of the curve is 5 microvolts per degree.

The Ta-Mo curve crosses the W-Ta curve at 1250°C or 17.3 millivolts. This temperature value agrees, as it should, with the temperature for zero output of the W-Mo couple. It reaches a maximum of 20.2 millivolts at 1700°C. Due to this maximum at such a low temperature, it has not been used for practical applications.

The W-Ta couple is being used for spectral emissivity determination of different materials. The couple is spot-welded to a 0.030-in. tungsten wire which is coated with the material being investigated. The W-Mo couple is being used for temperature measurements in a vacuum furnace, where optical pyrometry is difficult.

[5] B. Osann and E. Schroeder, Arch. f. d. Eisenhuttenwesen 7, 89-94 (1933).

Interaction of a Spiral Electron Beam and a Resonant Microwave Cavity*

F. K. Willenbrock and S. P. Cooke
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received July 18, 1949)

The interaction between the electromagnetic field in a cavity and an electron beam projected along the axis of the cavity is examined. The particular cavity considered here is of the cylindrical TE_{11n} type in a steady axial magnetic field. If the cavity is excited in a linearly polarized mode, the electromagnetic field will drive the electrons in a helical trajectory with an expanding radius, and the electrons will excite and transfer energy to a degenerate mode oriented spatially at right angles to the driving field. In the driving plane of polarization (both planes of polarization if the cavity is excited in a circularly polarized mode), the electron beam will excite a field in phase opposition to the driving field in a manner analogous to the counter e.m.f. in an electromechanical generator. The converse case of a TE_{11n} cavity excited by a spiral beam of electrons is also considered.

I. QUALITATIVE DISCUSSION OF SPIRALING ELECTRON BEAMS

THE behavior of a beam of electrons in a crossed electric and magnetic field has been widely discussed. In the recent literature,[1–3] several different uses of spiraling beams have been suggested. Malter[1] examines the impedance change in a parallel plate condenser through which an electron beam is projected in the presence of a steady magnetic field. Smith and Schulman[2] discuss the effect of spiraling electron beams on the frequency and Q of resonant systems. Spiral beam control has been used for frequency and amplitude modulation in several experimental magnetrons.[4–6]

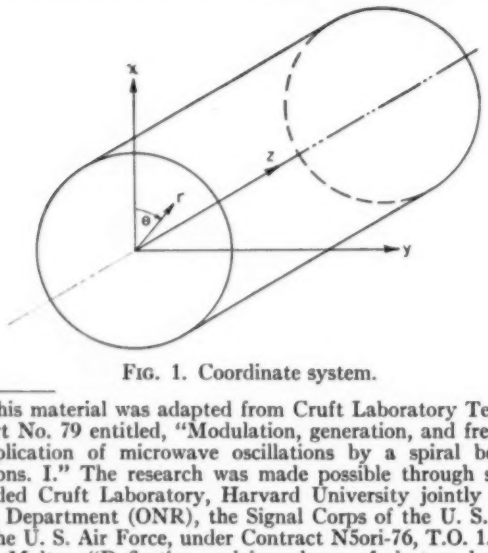


Fig. 1. Coordinate system.

Truell[3] considers the possibility of measuring the amplitude of an electromagnetic field by the deflection of an electron beam in the presence of a steady magnetic field.

The present investigation is a study of the interaction of an axial beam of electrons with a cylindrical cavity oscillating in a TE_{11n} mode with a steady magnetic field superimposed along the axis of the cavity. The theory of the interaction is presented both from the standpoint of energy transfer from the electromagnetic field to the electron beam and from the beam to the field.

To simplify the analysis, the trajectory of a single electron is considered first. A single spiraling electron, unlike an electron beam, does not cause a significant change of the r-f field in the cavity, and does not extract an appreciable part of the power driving the cavity. The electron follows a helical trajectory with an increasing radius in a cavity driven by a linearly or by a circularly polarized field. Linear polarization is obtained when the cavity is driven through a single coupling iris; circular polarization is obtained when the cavity is driven through two coupling holes located at right angles in space with equal input-signal amplitudes in time quadrature.

For a given cavity, the two parameters that can be readily varied are the magnetic field strength which determines the cyclotron angular frequency, $\omega_c = eB_0/m$, and the axial velocity of the electron beam, v_0, which determines the transit time of the electron through the cavity. It is found that for a cavity of length $\lambda_g/2$, maximum energy is transferred to the electron when $\omega_c = \omega_0$, the resonant angular frequency of the cavity. For a cavity of length λ_g with $\omega_0 = \omega_c$, all or nearly all the energy gained by the beam during the first half-wave-length will be returned to the r-f field during the second half-wave-length; and there will be a negligible amount of rotational energy in the beam at exit. The deflection of the electron varies inversely with v_0.

The treatment of the trajectory of a single electron in a cavity excited in a linearly polarized mode is also valid for an electron beam if (1) the effect of space charge is neglected; (2) the degenerate mode at right angles to the driving mode is suppressed; (3) the de-

* This material was adapted from Cruft Laboratory Technical Report No. 79 entitled, "Modulation, generation, and frequency multiplication of microwave oscillations by a spiral beam of electrons. I." The research was made possible through support extended Cruft Laboratory, Harvard University jointly by the Navy Department (ONR), the Signal Corps of the U. S. Army, and the U. S. Air Force, under Contract N5ori-76, T.O. 1.

[1] L. Malter, "Deflection and impedance of electron beams at high frequencies in the presence of a magnetic field," RCA Rev. **5**, 439–454 (April, 1941).

[2] L. P. Smith, and C. I. Schulman, "Frequency modulation and control of electron beams," Proc. I.R.E. **35**, 644–657 (July, 1947).

[3] R. Truell, "A method of measuring the field strength of high-frequency electromagnetic fields," Proc. I.R.E. **36**, 1249–1251 (October, 1948).

[4] Kilgore, Schulman, and Kurshan, "A frequency-modulated magnetron for super-high frequencies," Proc. I.R.E. **35**, 657–664 (July, 1947).

[5] Donal, Bush, Cuccia, and Hegbar, "A 1-Kilowatt Frequency-Modulated Magnetron for 900 Megacycles," Proc. I.R.E. **35**, 664–669 (July, 1947).

[6] J. S. Donal, Jr. and R. R. Bush, "A spiral-beam method for the amplitude modulation of magnetrons," Proc. I.R.E. **37**, 375–382 (April, 1949).

crease of the effective driving field to compensate for the energy transferred to the electron beam is taken into account.

The effect of space charge can be neglected because of the relatively low charge densities in the beam and the presence of the steady axial magnetic field. The beam in the cavity is not bunched as in a magnetron, but is a steady stream of electrons. Although the cross section of the beam is assumed to be small, the total beam currents considered are also small. For these reasons, high charge densities are not obtained. In any event, the presence of the strong axial magnetic field prevents defocusing of the beam.

Unless the perpendicular mode which is degenerate with the driving mode is suppressed, it will be excited by the spiraling electron beam. The r-f field in the cavity is, therefore, a superposition of a linear driving mode and a mode at right angles driven by the electron beam. This results in an elliptically polarized r-f field. If this degenerate mode is suppressed, the trajectory of the beam is the same as that of a single electron provided the decrease of effective driving field is considered.

The electron beam also excites a mode of oscillation in the original plane of polarization in phase opposition to the driving field. This "counter" E-field reduces the effective amplitude of the driving field and is analogous to the counter-e.m.f. in an electromechanical generator. With this reduced driving field, the losses in the cavity walls are also decreased. The remaining r-f input power is transferred by the beam to the other plane of polarization and appears as rotational energy in the beam at exit.

For the circularly polarized mode, there is a field driven through an iris by the external generator in each plane of polarization, and in each of these planes there are two counter-fields in phase opposition to the externally driven field. The first of these counter-fields is the same as that described for the linear case and is analogous to the counter-e.m.f. in an electromechanical generator. This is the "direct" counter-field which would exist even if the mode in the perpendicular plane were suppressed.

The second counter-field is cross-coupled by the electron beam from the driving field in the perpendicular plane of polarization. In the linear case this cross-coupled field drives the other degenerate mode and produces elliptical polarization. In the circular case, this cross-coupled field appears in opposition to the phase of the externally driven field if the latter has been chosen to add energy to the beam. The two counter-fields reduce the effective field more than the single counter-field in the linear case, and there is a correspondingly greater energy transfer to the beam for circular polarization.

If a spiraling beam of electrons is projected along the axis of the second TE_{11n_z} cavity, the cavity is excited, and energy is transferred from the beam to the r-f field. With the proper choice of parameters, it is possible to extract all the rotational energy from the beam. However, if the transit time is too long, the electron beam regains rotational energy from the r-f field in the cavity.

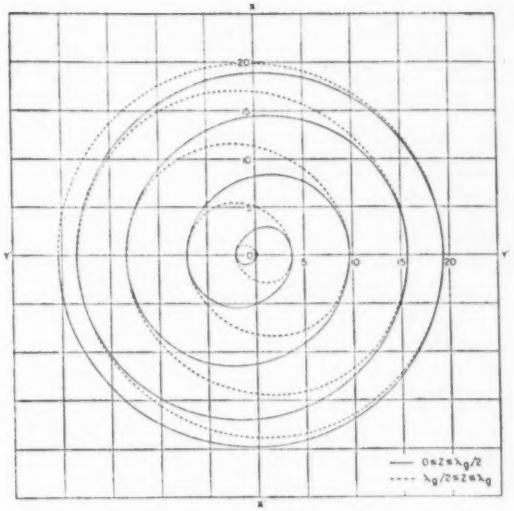


FIG. 2. Projection on the transverse plane of the trajectory of a single electron. TE_{11} cavity excited in a linearly polarized mode. $\kappa=1$, $\eta=0.1$, and $\phi=0$.

II. TRAJECTORY OF AN ELECTRON IN A TE_{11n_z} CAVITY IN A MAGNETIC FIELD

The transverse E-field in a perfectly conducting TE_{11n_z} cavity can be represented as $\mathbf{E}(r,\theta)\sin 2\pi z/\lambda_g \cos\omega_0 t$. For small deviations from the axis of the cavity $\mathbf{E}(r,\theta)$ can be written, by proper choice of the x axis, as $E_x=E_{0x}$, $E_y=0$, where E_{0x} is a constant. For velocities small in comparison with the velocity of light and for small r-f power inputs, the effect of the r-f magnetic field is negligible. The nonrelativistic equations of motion can be written† (see Fig. 1),

$$\left.\begin{aligned} \ddot{x} &= -e\mathcal{E}_x/m-(e/m)B_0\dot{y} \quad \text{(a)} \\ \ddot{y} &= -e\mathcal{E}_y/m+(e/m)B_0\dot{x} \quad \text{(b)} \\ \ddot{z} &= 0 \quad \text{(c)} \end{aligned}\right\} \tag{1}$$

where the dot notation indicates derivatives with respect to time, and e/m is the charge-to-mass ratio of the electron. The steady magnetic field, B_0, is directed along the positive z axis.

A linearly polarized mode as seen by the electron in its transit through the cavity can be written as

$$\left.\begin{aligned} \mathcal{E}_x &= E_{0x}\sin(2\pi z/\lambda_g)\cos(\omega_0 t+\phi) \\ \mathcal{E}_y &= \mathcal{E}_z = 0 \end{aligned}\right\} \tag{2}$$

for the interval $0\le t\le t_1$. Here, λ_g = wave-length in the cavity, $f_0=\omega_0/2\pi$ = resonant frequency of the cavity, t_1 = transit time of the electron through the cavity, and ϕ = phase angle of E-field at time of entrance.

† Rationalized M.K.S. units are used in this report.

The solution of Eq. (1c) is simply $z=v_0t$ where v_0 is the longitudinal velocity of the electron, and $z=0$ at $t=0$. Substituting this expression in the above equations will lead to solutions for $x(t)$ and $y(t)$ in terms of the parameters $\kappa=\omega_0/\omega_c$ and $\eta=v_0/v_p$, where $\omega_c=(e/m)B_0$ is the cylotron angular frequency, and v_p is the phase velocity of the r-f wave in the cavity. The solutions are for $\kappa=1$ (see Appendix A),

$$\left.\begin{aligned} x(t)=-\frac{eE_{0x}}{2m\omega_0^2}\Bigg\{&\frac{\sin[(1+\eta)\omega_0t+\phi]}{\eta(2+\eta)}-\frac{\sin[(1-\eta)\omega_0t+\phi]}{\eta(2-\eta)}+\frac{2}{\eta(4-\eta^2)}\\ &\times[(2-\eta^2)\sin\omega_0t\cos\phi+2\cos\omega_0t\sin\phi]\Bigg\}\\ y(t)=-\frac{eE_{0x}}{2m\omega_0^2}\Bigg\{&-\frac{2\eta\cos}{1-\eta^2}+\frac{\cos[(1+\eta)\omega_0t+\phi]}{\eta(2+\eta)(1+\eta)}+\frac{\cos[(1-\eta)\omega_0t+\phi]}{\eta(2-\eta)(1-\eta)}-\frac{2}{\eta(4-\eta^2)}\\ &\times[(2-\eta^2)\cos\omega_0t\cos\phi-2\sin\omega_0t\sin\phi]\Bigg\}\end{aligned}\right\}. \quad (3)$$

The electron traces a trajectory in the form of a helix with an expanding radius. Figure 2 shows a projection on the transverse (x, y) plane of the trajectory of the electron as it travels longitudinally through the cavity. For the chosen direction of the B-field, the electron spirals in a clockwise sense. The axes of the figures are

$$x'(t)=x(t)\Big/-\frac{eE_{0x}}{2m\omega_0^2}$$

and

$$y'(t)=y(t)\Big/-\frac{eE_{0x}}{2m\omega_0^2}.$$

This curve is plotted for $\eta=0.1$, which corresponds approximately to a 4000-volt electron. This high value of voltage was chosen to facilitate the calculation and plotting. For a more efficient transfer of energy to the beam, accelerating voltages of the order of 100 volts or less should be used.

Similarly, a circularly polarized mode as seen by the electron in the cavity can be written as

$$\left.\begin{aligned}\mathcal{E}_x&=E_{0x}\sin(2\pi/\lambda_g)z\cos\omega_0t\\ \mathcal{E}_y&=E_{0x}\sin(2\pi/\lambda_g)z\sin\omega_0t\\ \mathcal{E}_z&=0\end{aligned}\right\} \quad (4)$$

for the interval $0\leq t\leq t_1$. The direction of the B-field is taken the same as in the linearly polarized case. The circularly polarized field is chosen so that the E-field vector rotates in the clockwise sense. The solutions of the equations of motion for the same conditions are (see Appendix A),

$$\left.\begin{aligned} x(t)=-\frac{eE_{0x}}{2m\omega_0^2}\Bigg\{&\frac{2}{\eta}\sin\omega_0t-\frac{1}{\eta(1+\eta)}\sin(1+\eta)\omega_0t-\frac{1}{\eta(1-\eta)}\sin(1-\eta)\omega_0t\Bigg\}\\ y(t)=-\frac{eE_{0x}}{2m\omega_0^2}\Bigg\{&-\frac{2\eta}{1-\eta^2}-\frac{2}{\eta}\cos\omega_0t+\frac{1}{\eta(1+\eta)}\cos(1+\eta)\omega_0t+\frac{1}{\eta(1-\eta)}\cos(1-\eta)\omega_0t\Bigg\}\end{aligned}\right\}. \quad (5)$$

Figure 3, shows the projection on the transverse plane of the trajectory of a single electron for this case.

For $\kappa=\omega_0/\omega_c=1$, the maximum transverse deflection is obtained for both the linearly and circularly polarized fields. For a cavity of length λ_g, the electron gains energy during the first half-wave-length and loses it during the second half-wave-length.‡ If the transit time is chosen so that $1/\eta$ is an integer, all the rotational energy is lost during the second half wave-length. If $1/\eta$ is not an integer, there is a small amount of rotational energy left in the beam at exit. The deflection of the beam at $z=\lambda_g$ is of the order $1/\eta^2$ smaller than

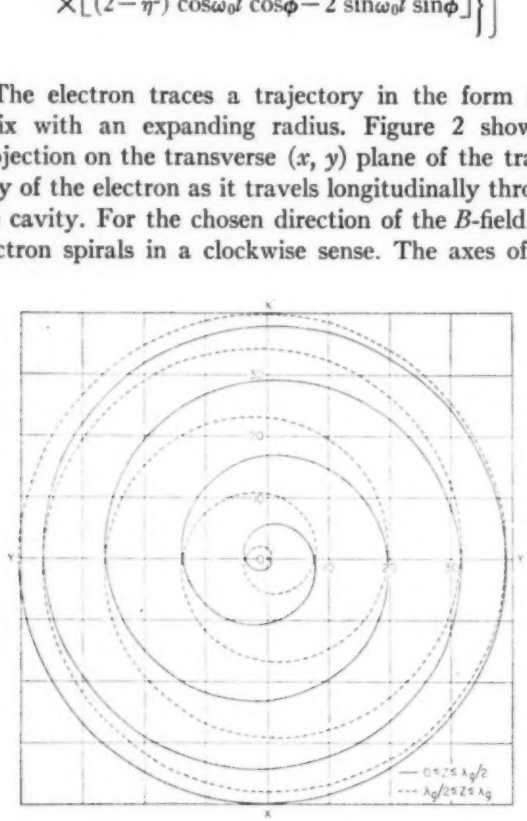


FIG. 3. Projection on the transverse plane of the trajectory of a single electron. TE_{11} cavity excited in a circularly polarized mode. $\kappa=1$, $\eta=0.1$.

‡ There are two unique cases, $\kappa=1/(1\pm\eta)$, for which there are equal increases in transverse deflection for each half wave-length traveled until relativistic effects become significant.

the deflection at $z=\lambda_g/2$. The deflection varies approximately as $1/\eta$. The transverse deflection for the circularly polarized case is approximately twice as great as the deflection for the linearly polarized case. For the same value of E_{0x}, it would require twice as much input power to drive the cavity in circular as in linear polarization.

For the $\kappa=1$ case, there are three sinusoidal terms in the expressions for $x(t)$ and $y(t)$ with frequencies $(1+\eta)\omega_0$, $(1-\eta)\omega_0$, and ω_0. The transit time, t_1, for a length $\lambda_g/2$ can be written as $\pi/\eta\omega_0$. Then, at time t_1 the arguments of these sinusoidal terms become

$$\begin{aligned}(1+\eta)\omega_0 t_1&=\pi/\eta+\pi\\(1-\eta)\omega_0 t_1&=\pi/\eta-\pi\\\omega_0 t_1&=\pi/\eta.\end{aligned}$$

The first two terms, $(1+\eta)\omega_0$ and $(1-\eta)\omega_0$, which describe the time dependence of the r-f driving E-field, differ by π-radians from their phase at $t=0$. Energy is gained by the electron in its transit from $z=0$ to $z=\lambda_g/2$ but is returned to the r-f field in the transit from $z=\lambda_g/2$ to $z=\lambda_g$.

By calculating the square root of the sum of x^2 and y^2 as given by Eq. (3), an expression is obtained for the radial deflection of the electron in terms of the entrance phase angle, ϕ, of the linearly polarized field. It is found that the terms independent of ϕ are of the order of $1/\eta^2$ larger than those containing ϕ. For values of $\eta\leq 0.1$, which are of the greatest practical interest, the radial deflection is effectively independent of ϕ. There is also a small spatial deviation of phase: while two electrons entering the cavity 90 degrees out-of-phase in time will still be 90 degrees out-of-phase in time at exit, they will not be 90 degrees out-of-phase in space. This bunching is negligible for values of $\eta\leq 0.1$.

For a circularly polarized mode, each electron has the same trajectory through the cavity except for the rotation of the coordinate system about the cylindrical axis.

In the solutions of the equations of motion for a single electron, the following approximations were made:

1. The E-field magnitude is constant for small deviations from the axis of the cavity.
2. The effect of the r-f B-field is negligible for velocities small with respect to the velocity of light and for small r-f power inputs.
3. The relativistic increase in electron mass is negligible.

For the choice of axes specified in Eqs. (2) for the linear mode, the expressions for the transverse E-field in a perfectly conducting TE_{11n_z} cavity are

$$\left.\begin{aligned}E_r&=-A_0(1/k_c r)J_1(k_c r)\cos\theta\\E_\theta&=A_0 J_1'(k_c r)\sin\theta\end{aligned}\right\},$$

where J_1=first-order Bessel function of first kind, J_1'=first derivative of J_1, $k_c=2\pi/\lambda_c$ where λ_c is the cut-off wave-length of the cavity, $A_0=2E_{0x}$=amplitude constant.

Using the coordinate system of Fig. 1, the E-field can be written as

$$\left.\begin{aligned}E_x&=A_0\{[+(1/k_c r)J_1(k_c r)\\&\qquad -\tfrac{1}{2}J_0(k_c r)]\cos 2\theta+\tfrac{1}{2}J_0(k_c r)\}\\E_y&=A_0[-(1/k_c r)J_1(k_c r)+\tfrac{1}{2}J_0(k_c r)]\sin 2\theta.\end{aligned}\right\}$$

For small values of the argument $(k_c r)$,

$$(1/k_c r)J_1(k_c r)-\tfrac{1}{2}J_0(k_c r)\doteq 0$$

and these equations reduce to

$$\begin{aligned}E_x&\doteq\tfrac{1}{2}A_0 J_0(k_c r)\doteq\tfrac{1}{2}A_0\\E_y&\doteq 0.\end{aligned}$$

This approximation is valid to within 1.2 percent for values of the argument up to 0.3. Values of the argument up to 0.6 introduce errors of less than 5 percent. In a cavity designed to resonate at 9000 mc/sec., an argument of 0.5 corresponds to a deflection of one-quarter of the radius of the cavity, and, for $\omega_0=\omega_c$, corresponds to a tangential velocity of the electron roughly equal to the velocity of light.

The second approximation depends on the velocity of the electron and the r-f power input to the cavity. If the ratio of the transverse velocity of the electron is small with respect to c, the velocity of light, the force on the electron exerted by the E-field is large with respect to the force exerted by the B-field. If the power input to the cavity is large (above a kilowatt for an X band cavity), the magnitude of the r-f B-field approaches the magnitude of the steady B-field and therefore becomes appreciable.

The relativistic increase of mass is negligible if the ratio of the total velocity of the electron to the velocity of light is small with respect to one. For electrons of less than 5000 volts energy, relativistic effects introduce an error of less than 1 percent.

III. TRAJECTORY OF AN ELECTRON BEAM IN A TE_{11n_z} CAVITY IN A MAGNETIC FIELD

The theoretical treatment of the trajectory of a single electron in a cavity must be modified when a beam of electrons is considered. If the cavity is cylindrically symmetrical, two degenerate TE_{11n_z} modes can exist having mutually perpendicular orientations. The spiraling electron beam serves to couple energy between these two modes, and this coupling must be taken into account in the calculation of the beam trajectory.

The azimuthal position angle, θ, of the electrons in the cavity is related to the deflections by $\tan\theta=y/x$. For linear polarization with $\kappa=1$ and $\eta\ll 1$, the approximate expression for θ (after several r-f cycles) is

$$\theta\doteq\omega_c t+\phi+\pi/2.$$

In this expression each electron has its own local time which is taken to be zero when the electron enters the cavity.

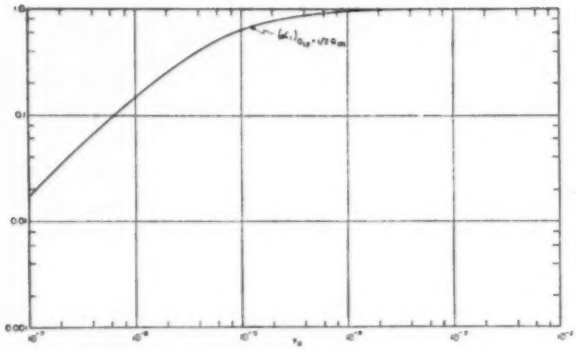


FIG. 4. $\alpha_1 = E_{2yz}/E_{1x}$ as a function of Y_0.
X band TE_{111} cavity with $\kappa = 1$.

If electron 1 enters at $\phi = 0$, its azimuthal angle (θ_1) is, after a time interval Δt_1, $\theta_1 = \omega_c \Delta t_1$. If electron 2 enters at $\phi = \phi_2$, its azimuthal angle (θ_2) is, after a time interval Δt_2, $\theta_2 = \omega_c \Delta t_2 + \phi_2$. Since $\phi_2 = \omega_c \Delta t_1 - \omega_c \Delta t_2$, $\theta_1 = \theta_2$. A projection on the transverse plane at any instant of a succession of electrons traversing the cavity would show a line of electrons each having the same azimuthal angle, but with greater radial deflections for the electrons which had been in the cavity for a longer interval of time. Accordingly, the beam of electrons has the form of a straight pencil at an angle to the axis of symmetry of the cavity rotating with an angular frequency ω_c.††

For circular polarization with $\kappa = 1$ and $\eta \ll 1$, the successive electrons will have the same azimuthal angle, as in the linear case. The beam has the same form of a straight pencil rotating around the axis of symmetry with an angular frequency ω_c.

For linear polarization, the spiraling electron beam drives a counter E-field in phase opposition to the original driving field. If the original r-f driving field is polarized in the plane containing the x axis, its amplitude E_{0x} is reduced by the counter E-field. This counter E-field is called E_{2xx} since it is directed along the x axis and is produced by the driving field in the plane of the x axis. With the reduction in the effective driving field, the losses in the cavity walls are decreased and the balance of the r-f input power appears in the rotational energy of the beam. The amplitude of the effective or reduced field is $|E_{1x}| = |E_{0x}| - |E_{2xx}|$.

The spiraling beam also excites the degenerate mode in the plane of the y axis. The amplitude of this cross-coupled field is called E_{2yz}. It is directed along the y axis, but is caused by the driving field in the plane of the x axis.

The differential equations of motion (1) for the electrons in the beam are unchanged, assuming the same approximations. For $\kappa = 1$ and $\eta^2 \ll 1$, the r-f field in the cavity as seen by the electron beam can be written as

$$\left.\begin{aligned} \mathcal{E}_{1x} &= E_{1x} \sin(2\pi z/\lambda_g) \cos\omega_0 t \\ \mathcal{E}_{2yz} &= -E_{2yz} \sin(2\pi z/\lambda_g) \sin\omega_0 t \\ \mathcal{E}_z &= 0 \end{aligned}\right\} \tag{6}$$

for the interval $0 \leq t \leq t_1$. The time dependence chosen for the E_{2yz} field is valid since the electron beam is a straight pencil beam rotating around the axis of symmetry. The phase angle of the E_{1x} field at entrance is omitted since $\eta^2 \ll 1$. The solution of the equations of motion is for $\kappa = 1$ (see Appendix A),

$$\left.\begin{aligned} x(t) = -\frac{eE_{1x}}{2m\omega_0^2}\Bigg[& \frac{4-2\eta^2}{\eta(4-\eta^2)} \sin\omega_0 t \\ & - \frac{\sin(1+\eta)\omega_0 t}{\eta(2+\eta)} - \frac{\sin(1-\eta)\omega_0 t}{\eta(2-\eta)} \\ & + \alpha_1 \Bigg\{ \frac{-4}{\eta(4-\eta^2)} \sin\omega_0 t + \frac{\sin(1+\eta)\omega_0 t}{\eta(2+\eta)(1+\eta)} \\ & \qquad + \frac{\sin(1-\eta)\omega_0 t}{\eta(2-\eta)(1-\eta)} \Bigg\} \Bigg] \\ y(t) = -\frac{eE_{1x}}{2m\omega_0^2}\Bigg[& \frac{-2\eta}{1-\eta^2} - \frac{4-2\eta^2}{\eta(4-\eta^2)} \cos\omega_0 t \\ & + \frac{\cos(1+\eta)\omega_0 t}{\eta(2+\eta)(1+\eta)} + \frac{\cos(1-\eta)\omega_0 t}{\eta(2-\eta)(1-\eta)} \\ & + \alpha_1 \Bigg\{ \frac{4}{\eta(4-\eta^2)} \cos\omega_0 t - \frac{\cos(1+\eta)\omega_0 t}{\eta(2+\eta)} \\ & \qquad - \frac{\cos(1-\eta)\omega_0 t}{\eta(2-\eta)} \Bigg\} \Bigg] \end{aligned}\right\} \tag{7}$$

This solution is written in terms of α_1 which is defined as the ratio of the E-field amplitudes, E_{2yz}/E_{1x}.

For a cavity excited in a circularly polarized mode, it is necessary to introduce E-field amplitudes for each plane of polarization corresponding to those defined for the linear case. The original driving fields in each plane have the amplitudes E_{0x} and E_{0y} when no energy is transferred to the beam. The reduced amplitudes are E_{1x} and E_{1y}.

There are two counter-fields in each plane: The first is the counter-field produced by the driving field in the same plane; the second is the counter-field cross-coupled from the driving field in the perpendicular plane. The first counter-field is analogous to the counter-e.m.f. in an electromechanical generator. The amplitude of this field is E_{2xx} in the x axis plane and E_{2yy} in the y axis plane. It acts in the same plane as the driving field by which it is produced.

The second type of counter-field, the cross-coupled field, acts in the plane perpendicular to the driving

†† A graphical analysis indicates that the $\eta \ll 1$ restriction is stronger then necessary and that a pencil beam is formed for $\eta^2 \ll 1$.

field by which it is produced. The field E_{2yz} acts in the y axis plane, but is caused by the driving field in the x axis plane. The field E_{2xy} acts in the x axis plane, but is caused by the driving field in the y axis plane.

Both types of counter-fields act in opposition to the phase of the driving fields if these original fields are chosen to add energy to the beam. Therefore

$$\begin{aligned} |E_{1x}| &= |E_{0x}| - |E_{2xz}| - |E_{2xy}| \\ |E_{1y}| &= |E_{0y}| - |E_{2yy}| - |E_{2yz}| . \end{aligned} \tag{8}$$

The differential equations of motion (1) are unchanged assuming the same approximations. For $\kappa=1$ and $\eta \ll 1$, the r-f field in the cavity as seen by the electron beam is

$$\left.\begin{aligned} \mathcal{E}_{1x} &= E_{1x} \sin(2\pi z/\lambda_g) \cos\omega_0 t \\ \mathcal{E}_{1y} &= E_{1x} \sin(2\pi z/\lambda_g) \sin\omega_0 t \\ \mathcal{E}_{1z} &= 0 \end{aligned}\right\} \tag{9}$$

for the interval $0 \leq t \leq t_1$. Except for the change in amplitude these expressions are identical with those for a single electron (4). Therefore, the solutions for a single electron (5) will be valid for the electron beam if E_{1x} is substituted for E_{0x}.

IV. TRAJECTORY OF A SPIRAL BEAM EXCITING A TE_{11n_z} CAVITY IN A MAGNETIC FIELD

If a spiraling beam of electrons is projected into a TE_{11n_z} cavity in the presence of a steady axial magnetic field, the cavity is excited, and energy is extracted from the electron beam. It is assumed in this discussion that a steady state condition has been reached. It is also assumed that the spiraling beam is a straight pencil of electrons rotating about the axis of symmetry at an angular frequency ω_c. The phase of the E-field in the cavity is determined by the phase of the electrons upon entrance ($t=0$). If the phase of the entering beam is chosen so that $x_0=0$, $\dot{y}_0=0$, $\dot{x}_0$ has its maximum negative value, and y_0 has its maximum positive value, then the r-f field as seen by the electron beam for $\kappa=1$ and $\eta^2 \ll 1$ is

$$\left.\begin{aligned} \mathcal{E}_{2x} &= -E_{2x} \sin(2\pi z/\lambda_g) \cos\omega_0 t \\ \mathcal{E}_{2y} &= -E_{2y} \sin(2\pi z/\lambda_g) \sin\omega_0 t \\ \mathcal{E}_{2z} &= 0 \end{aligned}\right\} \tag{10}$$

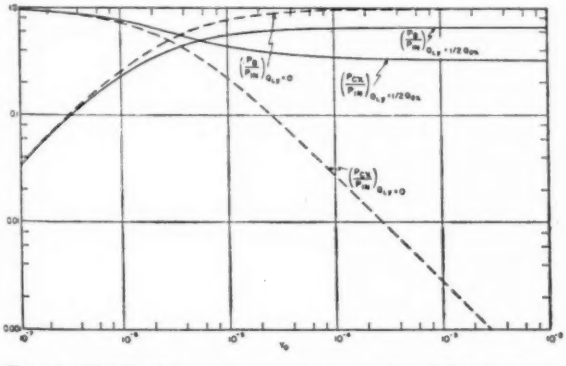


FIG. 5. Distribution of power between the driving plane of polarization and the electron beam. X band TE_{111} cavity with $\kappa=1$.

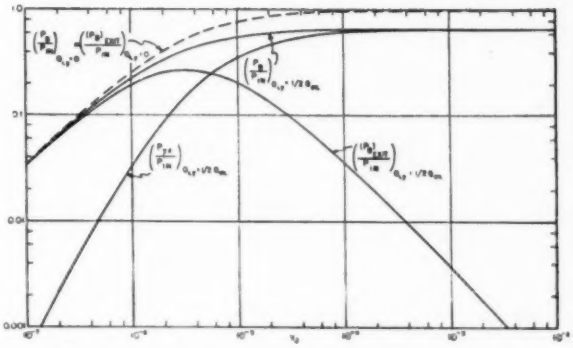


FIG. 6. Distribution of power between the electron beam and the driven plane of polarization. X band TE_{111} cavity with $\kappa=1$.

for the interval $0 \leq t \leq t_1$. The trajectory of the beam can be obtained from the solution of the differential equations of motion (1). The solution is for $\kappa=1$ (see Appendix B),

$$\left.\begin{aligned} x(t) = {} & \frac{\dot{x}_0}{\omega_0} \sin\omega_0 t + \frac{eE_{2x}}{2m\omega_0^2} \left\{ \frac{4-2\eta^2}{\eta(4-\eta^2)} \sin\omega_0 t \right. \\ & \left. - \frac{\sin(1+\eta)\omega_0 t}{\eta(2+\eta)} - \frac{\sin(1-\eta)\omega_0 t}{\eta(2-\eta)} \right\} \\ & + \frac{eE_{2y}}{2m\omega_0^2} \left\{ \frac{4}{\eta(4-\eta^2)} \sin\omega_0 t \right. \\ & \left. - \frac{\sin(1+\eta)\omega_0 t}{\eta(1+\eta)(2+\eta)} - \frac{\sin(1-\eta)\omega_0 t}{\eta(1-\eta)(2-\eta)} \right\} \\ y(t) = {} & y_0 + \frac{\dot{x}_0}{\omega_0} - \frac{\dot{x}_0}{\omega_0} \cos\omega_0 t \\ & + \frac{eE_{2x}}{2m\omega_0^2} \left\{ -\frac{2\eta}{1-\eta^2} - \frac{4-2\eta^2}{\eta(4-\eta^2)} \cos\omega_0 t \right. \\ & \left. + \frac{\cos(1+\eta)\omega_0 t}{\eta(1+\eta)(2+\eta)} + \frac{\cos(1-\eta)\omega_0 t}{\eta(1-\eta)(2-\eta)} \right\} \\ & + \frac{eE_{2y}}{2m\omega_0^2} \left\{ \frac{-4}{\eta(4-\eta^2)} \cos\omega_0 t \right. \\ & \left. + \frac{\cos(1+\eta)\omega_0 t}{\eta(2+\eta)} + \frac{\cos(1-\eta)\omega_0 t}{\eta(2-\eta)} \right\} \end{aligned}\right\} . \tag{11}$$

V. COUPLING BETWEEN OSCILLATIONS IN MUTUALLY PERPENDICULAR PLANES OF POLARIZATION IN A CAVITY DRIVEN IN LINEAR POLARIZATION

The power input (P_{in}) to the cavity excited in linear polarization in the x axis plane is partially dissipated in the walls of the cavity (P_{cx}) and partially transferred to the electron beam (P_B). P_B is determined from

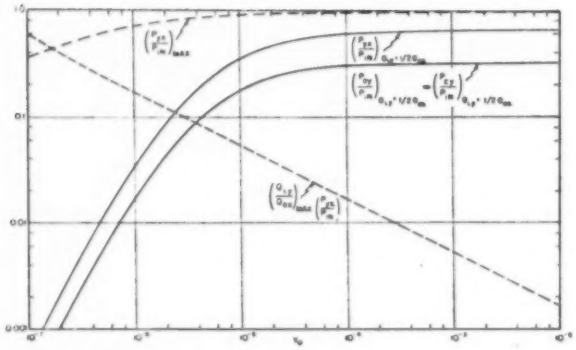


FIG. 7. Distribution of power in the plane of polarization driven by the electron beam. X band TE_{111} cavity with $\kappa=1$.

the beam current and the driving field along the trajectory of the beam,

$$P_B=\int \text{beam current} \times \text{driving field}.$$

The trajectory of the beam of electrons in a linearly excited cavity is given in Section III in terms of E_{1z} and $\alpha_1=E_{2yz}/E_{1z}$.

P_{cz} is obtained with the usual expression for Q,

$$Q=\omega_0\frac{\text{Energy stored}}{\text{Average power loss}},$$

in terms of P_{in}, E_{1z}, and the properties of the cavity. From these two expressions, P_B and P_{cz} can be written in terms of α_1 and the physical constants of the cavity.

To find α_1, the power coupled from the electron beam to the second plane of polarization, P_{yz}, must be evaluated. P_{yz} is found by taking the line integral of the product of the beam current and the driven E-field, E_{2yz}, along the trajectory of the beam. With the expression for the Q in the driven plane, another relation between E_{2yz} and P_{yz} is available. These two expressions can be solved simultaneously to give α_1 in terms of the beam admittance, $Y_0=I_0/V_0$, and cavity constants. With α_1 determined, P_B, P_{cz}, P_{yz}, and the rotational power left in the beam at exit can be written in terms of the r-f power input, P_{in}. If there is an iris coupled to the second plane of polarization, the power transmitted through the iris (P_{0y}) and the power dissipated in the cavity walls in the y axis plane (P_{cy}) can also be determined.

In the following treatment, α_1 is obtained first. The general power expressions are specialized to the case of a cavity of length $\lambda_g/2$ with $\kappa=1$ and $\eta^2\ll1$. The resulting expressions are plotted for a 9000-mc/sec. (X band) TE_{111} cavity.

The power coupled to the plane containing the y axis is given by

$$P_{yz}=Ne\int \boldsymbol{E}_2\cdot\mathbf{ds}, \tag{12}$$

where $N=$number of electrons entering per second,

$\boldsymbol{E}_2=-y'E_{2yz}\sin2\pi z/\lambda_g\sin\omega_0 t$,

$\mathbf{ds}=\boldsymbol{x}\dot{x}dt+\boldsymbol{y}\dot{y}dt+\boldsymbol{z}\dot{z}dt=$path element of the trajectory.

Since both $\boldsymbol{\xi}_2$ and $\mathbf{s}$ can be written as functions of t only, using $z=v_0t$, the limits of integration can be taken as $t=0$ and $t=t_1$, where t_1 is the transit time of an electron through the cavity.

For the case of a cavity of length $\lambda_g/2$ with $\kappa=1$, expression (12) becomes with (7) since $\eta^2\ll1$,

$$P_{yz}=\frac{I_0eE_{2zy}^2}{2m\omega_0^2\eta^2}\frac{(1-\alpha_1)}{\alpha_1}=\frac{Y_0\lambda_g^2E_{2yz}^2}{16\pi^2}\frac{(1-\alpha_1)}{\alpha_1}, \tag{13}$$

where $I_0=Ne=$beam current, $Y_0=I_0/V_0=$beam admittance, and $V_0=mv_0^2/2e=$beam accelerating voltage. With the usual definition of the cavity Q,

$$\left.\begin{aligned}P_{yz}&=\omega_0\left(\tfrac{1}{2}\epsilon_0\int E_2^2d\tau/Q_{Ly}\right)\\&=\omega_0\epsilon_0E_{2yz}^2\tau_0/2Q_{Ly}\end{aligned}\right\}, \tag{14}$$

where $Q_{Ly}=$the y plane Q loaded by the coupling iris, $\epsilon_0=$dielectric constant of free space, $d\tau=$element of volume in the cavity, $\tau_0=$volume constant (see Appendix C). From expressions (13) and (14), the desired result is obtained:

$$\alpha_1=\frac{1}{1+(8\pi^2\omega_0\epsilon_0\tau_0)/(Y_0Q_{Ly}\lambda_g^2)}. \tag{15}$$

To relate P_{yz} to P_{in}, it is first necessary to evaluate the power coupled into the electron beam, P_B.

$$P_B=-Ne\int \boldsymbol{E}_1\cdot\mathbf{ds}, \tag{16}$$

where $E'=\boldsymbol{x}E_{1z}\sin(2\pi z/\lambda_g)\cos\omega_0t$. For the case of a cavity of length $\lambda_g/2$, with $\kappa=1$ since $\eta^2\ll1$,

$$P_B=\frac{I_0eE_{1z}^2}{2m\omega_0^2}\frac{1}{\eta^2}(1-\alpha_1)=\frac{Y_0\lambda_g^2E_{1z}^2}{16\pi^2}(1-\alpha_1). \tag{17}$$

The power driving the cavity, P_{in}, is partially dissipated in the walls of the cavity and the rest is converted into rotational energy of the electrons. The power dissipated in the walls of the cavity in the x axis plane, P_{cx}, can be written as

$$P_{cx}=\omega_0\epsilon_0E_{1z}^2\tau_0/2Q_{0x}, \tag{18}$$

where $Q_{0x}=$unloaded Q of the mode in the x axis plane. Defining k_1 as the ratio P_B/P_{cx},

$$k_1=P_B/P_{cx}=Y_0Q_{0x}\lambda_g^2(1-\alpha_1)/8\pi^2\omega_0\epsilon_0\tau_0 \tag{19}$$

for the case of $\kappa=1$ and a cavity length of $\lambda_g/2$. Sub-

stituting the expression for α_1, (15) results in

$$k_1=(Q_{0x}/Q_{Ly})\alpha_1. \tag{20}$$

Since $P_{in}=P_{cx}+P_B$,

$$\left.\begin{aligned} P_B&=\frac{k_1}{1+k_1}P_{in}=\frac{1}{1+(Q_{Ly})/(Q_{0x}\alpha_1)}P_{in} \\ \text{and} \\ P_{cx}&=\frac{1}{1+k_1}P_{in}=\frac{1}{1+(Q_{0x}/Q_{Ly})\alpha_1}P_{in} \end{aligned}\right\}. \tag{21}$$

By defining two more coupling coefficients

$$k_2=P_{yx}/P_B \quad \text{and} \quad k_3=P_{0y}/P_{yx},$$

where P_{0y} is the power coupled out through an iris in the y axis plane, it is possible to write P_{0y} in terms of P_{in}.

The coefficient k_3 depends on the properties of the coupling iris and can be written as

$$k_3=1-Q_{Ly}/Q_{0y}, \tag{22}$$

where Q_{0y} is the unloaded Q in the y axis plane. From expressions (13) and (17)

$$k_2=\alpha_1$$

and

$$\frac{P_{yx}}{P_{in}}=\frac{\alpha_1}{1+(Q_{Ly})/(Q_{0x}\alpha_1)}. \tag{23}$$

Then, by the definitions of the coupling coefficients

$$\frac{P_{0y}}{P_{in}}=\frac{k_1k_2k_3}{1+k_1}=\frac{\alpha_1}{1+(Q_{Ly})/(Q_{0x}\alpha_1)}\left(1-\frac{Q_{Ly}}{Q_{0y}}\right). \tag{24}$$

Thus, the ratio P_{0y}/P_{in} is specified in terms of the beam admittance and the physical characteristics of the cavity.

The rotational power remaining in the spiral beam at the cavity exit, $(P_B)_{exit}$, can be obtained by subtracting P_{yx} from P_B. It is also possible to obtain $(P_B)_{exit}$ from kinetic energy considerations:

$$\left.\begin{aligned} (P_B)_{exit}&=Nm/2[(\dot{x})^2{}_{t=t_1}+(\dot{y})^2{}_{t=t_1}] \\ &=\frac{I_0eE_{1x}{}^2}{2m\omega_0{}^2\eta^2}(1-\alpha_1)^2=\frac{1-\alpha_1}{1+(Q_{Ly})/(Q_{0x}\alpha_1)}P_{in} \end{aligned}\right\} \tag{25}$$

for a cavity of length $\lambda_g/2$ with $\kappa=1$ and $\eta^2\ll 1$.

The amplitudes of the E-fields defined in Section III can now be readily evaluated in terms of P_{in} for a cavity of length $\lambda_g/2$ with $\kappa=1$. By definition of Q_{0x},

$$E_{0x}=(2Q_{0x}P_{in}/\omega_0\epsilon_0\tau_0)^{\frac{1}{2}}. \tag{26}$$

From (18) and (21),

$$E_{1x}=\left[\frac{2Q_{0x}}{\omega_0\epsilon_0\tau_0}\frac{P_{in}}{1+(Q_{0x}/Q_{Ly})\alpha_1}\right]^{\frac{1}{2}}. \tag{27}$$

From (27) and the definition of α_1,

$$E_{2yx}=\left[\frac{2Q_{0x}}{\omega_0\epsilon_0\tau_0}\frac{\alpha_1{}^2P_{in}}{1+(Q_{0x})/(Q_{Ly})\alpha_1}\right]^{\frac{1}{2}}. \tag{28}$$

The "counter" E-field, E_{2xx} defined in Section III, is obtained from (26) and (27),

$$E_{2xx}=\left\{\frac{2Q_{0x}P_{in}}{\omega_0\epsilon_0\tau_0}\left[1-\frac{1}{1+(Q_{0x}/Q_{Ly})\alpha_1}\right]\right\}^{\frac{1}{2}}. \tag{29}$$

The actual field present in the cavity is elliptically polarized and rotates in the sense opposite to that of the electrons.

In order to plot some of the foregoing expressions as functions of Y_0, the following values are assumed for an X band TE_{111} cavity with two coupling holes:

$$\begin{aligned} \omega_0&=2\pi\times 9\times 10^9 \text{ c.p.s.}, \\ \lambda_g&=4\times 10^{-2} \text{ meter}, \\ Q_{0x}&=Q_{0y}=2\times 10^4, \\ Q_{Ly}&=10^4, \\ \tau_0&=2.28\times 10^{-6} \text{ meter}^3 \end{aligned}$$

(see Appendix C). All plots are for $\eta^2\ll 1$ and $\kappa=1$, which corresponds to a B-field of 0.32 weber/meter^2 (3200 gauss). A practical range of values for the beam admittance, Y_0, is 10^{-2} to 10^{-7}: $Y_0=10^{-2}$ mhos corresponds to currents of the order of 10 or 100 ma, and voltages of the order of 1 to 10 volts; $Y_0=10^{-7}$ mhos corresponds to currents of the order of 10μa, and voltages of the order of 100 volts.

Figure 4 shows a plot of $\alpha_1=E_{2yx}/E_{1x}$ as a function of Y_0, Eq. (15).

Figure 5 shows a plot of the distribution of power ($P_{in}=P_B+P_{cx}$) between the driving (x axis) plane of polarization and the electron beam as a function of Y_0.

If the mode in the y axis plane is suppressed ($Q_{Ly}=0$), the approximate expressions (21) become indeterminate and it is necessary to let α_1 vanish in the expression for P_B (17). Then

$$P_B=Y_0\lambda_g{}^2E_{1x}{}^2/16\pi^2. \tag{30}$$

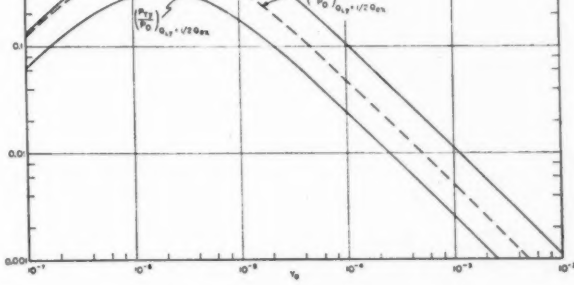


FIG. 8. Power transfer from a spiraling beam of electrons to an X band TE_{111} cavity. $\kappa=1$ and $Q_{Lx}\neq Q_{Ly}$.

Since P_{cx} remains the same,

$$\left.\begin{aligned}\frac{P_B}{P_{\text{in}}}&=\frac{1}{1+(8\pi^2\omega_0\epsilon_0\tau_0)/(Y_0\lambda_g^2Q_{0x})}\\ \text{and}\\ \frac{P_{cx}}{P_{\text{in}}}&=\frac{1}{1+(Y_0\lambda_g^2Q_{0x})/(8\pi^2\omega_0\epsilon_0\tau_0)}\end{aligned}\right\}. \tag{31}$$

These curves are also shown on Fig. 5.

Figure 6 shows the distribution of the power coupled into the beam ($P_B=P_{yx}+P_B$) as a function of Y_0. If the mode in the y axis plane is suppressed, ($Q_{Ly}=0$), $P_{yx}=0$, and $(P_B)_{\text{exit}}=P_B$.

Figure 7 shows the distribution of power in the driven (y axis) plane of polarization ($P_{yx}=P_{cy}+P_{0y}$) as a function of Y_0. P_{cy} is the power dissipated in the walls of the cavity by the oscillation in the y axis plane and can be written

$$\left.\begin{aligned}P_{cy}&=\frac{\omega_0\epsilon_0\tau_0E_{2yx}^2}{2Q_{0y}}\\ &=\frac{\alpha_1}{1+(Q_{Ly})/(Q_{0x}\alpha_1)}\frac{Q_{Ly}}{Q_{0y}}P_{\text{in}}\end{aligned}\right\} \tag{32}$$

by making use of (28). Since critical coupling ($Q_{0y}=2Q_{Ly}$) was chosen, $P_{0y}=P_{cy}$. It is of interest to determine the dependence of P_{yx} as a function of Q_{Ly}. Using expression (15) for α_1 and differentiating the power ratio P_{yx}/P_{in} (23), with respect to Q_{Ly}/Q_{0x}, the following condition results for maximum P_{yx}/P_{in},

$$\frac{Q_{Ly}}{Q_{0x}}=\left[\frac{8\pi^2\omega_0\epsilon_0\tau_0}{\lambda_g^2Y_0Q_{0x}}+\left(\frac{8\pi^2\omega_0\epsilon_0\tau_0}{\lambda_g^2Y_0Q_{0x}}\right)^2\right]^{\frac{1}{2}}. \tag{33}$$

Substituting this expression in (23), the optimum power ratio P_{yx}/P_{in} is obtained. This optimum power ratio and the ratio Q_{Ly}/Q_{0x} are also plotted in Fig. 7.

The value of Q_{Ly} can be decreased by increasing the size of the coupling hole in the y axis plane. Since it is not possible to enlarge the hole very much beyond critical coupling without materially decreasing the values of Q_{0y} and Q_{0x}, the maximization condition (33) has a limited validity.

VI. COUPLING BETWEEN OSCILLATIONS IN MUTUALLY PERPENDICULAR PLANES OF POLARIZATION IN A CAVITY DRIVEN IN CIRCULAR POLARIZATION

Since all the power in the beam appears at exit for the case of circular polarization, $P_B=(P_B)_{\text{exit}}$. From kinetic energy considerations,

$$P_B=Nm/2[(\dot{x})^2{}_{t=t_1}+(\dot{y})^2{}_{t=t_1}],$$

where $\dot{x}$ and $\dot{y}$ are found by differentiating expressions (5) with E_{1x} substituted for E_{0x}. For a cavity of length $\lambda_g/2$ with $\kappa=1$ for $\eta\ll1$,

$$P_B=(Y_0E_{1x}^2\lambda_g^2)/4\pi^2. \tag{34}$$

For a cylindrically symmetrical cavity, one-half of the power in the beam is coupled from the driving field in each of the two planes of polarization. Thus it is possible to define coefficients, similar to k_1 (19),

$$k_{1x}=P_{Bx}/P_{cx}$$

and

$$k_{1y}=P_{By}/P_{cy},$$

where $P_{Bx}=P_{By}=\frac{1}{2}P_B$. Then $k_{1x}=k_{1y}$. Using (18) for P_{cx}

$$k_{1x}=\frac{Y_0\lambda_g^2Q_{0x}}{4\pi^2\omega_0\epsilon_0\tau_0} \tag{35}$$

is obtained. By analogy with the expressions (21),

$$\left.\begin{aligned}P_{Bx}&=[k_{1x}/(1+k_{1x})](P_{\text{in}})_x\\ P_{cx}&=[1/(1+k_{1x})](P_{\text{in}})_x\end{aligned}\right\}. \tag{36}$$

Similar expressions hold for the y axis plane.

From (36) and (18), E_{1x} can be written explicitly as

$$E_{1x}=\left[\frac{2Q_{0x}}{\omega_0\epsilon_0\tau_0}\frac{(P_{\text{in}})_x}{1+k_{1x}}\right]^{\frac{1}{2}} \tag{37}$$

and by definition

$$E_{0x}=\left[\frac{2Q_{0x}}{\omega_0\epsilon_0\tau_0}(P_{\text{in}})_x\right]^{\frac{1}{2}}.$$

From expression (8),

$$E_{2xx}+E_{2xy}=E_{0x}-E_{1x}=\left[\frac{2Q_{0x}(P_{\text{in}})_x}{\omega_0\epsilon_0\tau_0}\left(\frac{k_{1x}}{1+k_{1x}}\right)\right]^{\frac{1}{2}}. \tag{38}$$

Similar expressions hold for the y axis plane.

VII. EXCITATION OF A TE$_{11n_z}$ CAVITY BY A SPIRALING BEAM OF ELECTRONS

Since the transient case is of little practical interest, it is assumed that steady-state conditions have been reached.

In Section IV, the expressions for the trajectory of the electron beam were determined for the case of a cavity excited by a spiraling beam. The power transferred to the cavity is found by taking the line integral of the product of the beam current and the E-field along the trajectory of the beam through the cavity. If the Q's of the two modes are not the same, it is convenient to evaluate separately the line integrals of the x and y components of the r-f field. The general expressions are specialized to the case of a cavity of length $\lambda_g/2$ with $\kappa=1$ and $\eta^2\ll1$.

The power transferred to the cavity, P_T, is given by the integral

$$P_T = +Ne\int \boldsymbol{E}_2 \cdot \mathbf{ds}, \tag{39}$$

where

$$\boldsymbol{E}_2 = -\boldsymbol{x}E_{2x}\sin(2\pi z/\lambda_g)\cos\omega_0 t - \boldsymbol{y}E_{2y}\sin(2\pi z/\lambda_g)\sin\omega_0 t$$

and

$$\mathbf{ds} = \boldsymbol{x}\dot{x}dt + \boldsymbol{y}\dot{y}dt.$$

Considering first the case where $Q_{Lx} \neq Q_{Ly}$, it is convenient to write P_T as the sum $(P_{Tx}+P_{Ty})$ of the powers transferred to the two modes. Then

$$\left.\begin{aligned} P_{Tx} &= -Ne\int \boldsymbol{x}E_{2x}\cdot\mathbf{ds} \\ \text{and} \\ P_{Ty} &= -Ne\int \boldsymbol{y}E_{2y}\cdot\mathbf{ds} \end{aligned}\right\}. \tag{40}$$

The evaluation of the first of these integrals for a cavity of length $\lambda_g/2$ with $\kappa=1$ results in since $\eta^2 \ll 1$,

$$P_{Tx} = -\frac{NeE_{2x}}{\omega_0\eta}\dot{x}_0 - \frac{Ne^2E_{2x}^2}{2m\omega_0^2\eta^2} - \frac{Ne^2E_{2x}E_{2y}}{2m\omega_0^2\eta^2}. \tag{41}$$

With the usual expression for the cavity Q,

$$P_{Tx} = \omega_0\epsilon_0\tau_0 E_{2x}^2/2Q_{Lx}. \tag{42}$$

These expressions determine E_{2x},

$$E_{2x} = \frac{-2I_0\dot{x}_0/\omega_0\eta}{(\omega_0\epsilon_0\tau_0/Q_{Lx}) + (Y_0\lambda_g^2/8\pi^2)(1+Q_{Ly}/Q_{Lx})} \tag{43}$$

where $\dot{x}_0$ is negative. Using (42)

$$P_{Tx} = \frac{\omega_0\epsilon_0\tau_0}{2Q_{Lx}} \times\left[\frac{-2I_0\dot{x}_0/\omega_0\eta}{(\omega_0\epsilon_0\tau_0/Q_{Lx}) + (Y_0\lambda_g^2/8\pi^2)(1+Q_{Ly}/Q_{Lx})}\right]^2. \tag{44}$$

Defining the driving power as

$$P_D = (Nm/2)\dot{x}_0^2, \tag{45}$$

it is possible to write (44) as the power ratio,

$$\frac{P_{Tx}}{P_D} = \frac{Q_{Lx}Y_0\lambda_g^2}{2\pi^2\omega_0\epsilon_0\tau_0} \times\left[\frac{1}{1+[Y_0\lambda_g^2(Q_{Lx}+Q_{Ly})]/[8\pi^2\omega_0\epsilon_0\tau_0]}\right]^2. \tag{46}$$

Similar expressions hold for the y axis plane.

If the cavity has one iris which couples to the y axis plane, Q_{Lx} will be greater than Q_{Ly}, and more power is coupled to the unloaded plane of polarization.

For the special case where $Q_{Lx}=Q_{Ly}$, $E_{2x}=E_{2y}$ and the total power transferred is $2P_{Tx}$,

$$\frac{P_T}{P_D} = \frac{Q_{Lx}Y_0\lambda_g^2}{\pi^2\omega_0\epsilon_0\tau_0}\left[\frac{1}{1+[Q_{Lx}Y_0\lambda_g^2]/[4\pi^2\omega_0\epsilon_0\tau_0]}\right]^2. \tag{47}$$

Figure 8 shows the power ratio P_T/P_D (47) as a function of Y_0 for a cylindrically symmetrical X band cavity having the same physical constants as used in Section V with Q_{Lx} taken equal to Q_{Ly}. Above the critical value of Y_0 for which P_T/P_D equals unity, the power transferred to the cavity drops off rapidly. The physical explanation is that for higher values of Y_0, the electrons lose all their rotational energy to the r-f field before they complete their transit through the cavity, and start to regain rotational energy from the field.

Figure 8 also shows P_{Tx}/P_D (46) and P_{Ty}/P_D (46) plotted as functions of Y_0 for the case where there is a single iris critically coupled to the y axis plane. Then

$$Q_{Lx} = Q_{0x} = 2Q_{Ly}.$$

The total power transferred to the beam $(P_{Tx}+P_{Ty})$ is

$$\frac{P_T}{P_D} = (Q_{Lx}+Q_{Ly})\frac{Y_0\lambda_g^2}{2\pi^2\omega_0\epsilon_0\tau_0} \times\left[\frac{1}{1+[Y_0\lambda_g^2(Q_{Lx}+Q_{Ly})]/[8\pi^2\omega_0\epsilon_0\tau_0]}\right]^2.$$

Maximizing this ratio with respect to Y_0 determines the critical value of Y_0,

$$Y_0 = \frac{8\pi^2\epsilon_0\omega_0\tau_0}{\lambda_g^2(Q_{Lx}+Q_{Ly})}.$$

VIII. CONCLUSIONS

Several possible applications of the interaction between the resonant cavity oscillations and the spiral beam of electrons are evident. First, this interaction could be used as the basic energy interchange mechanism in a microwave generator. Second, the coupling between oscillations in the two planes of polarization in a single cavity or the coupling between two cavities by an electron beam could serve as a method of amplitude modulating a microwave signal. Modulation could be effected by varying either the beam current or voltage. A tube called the Electron Coupler, which uses a two-cavity system for amplitude modulation, has

recently been described in the literature.[7] Third, the resonant cavity driven by an external oscillator could be used as a microwave circular sweep generator. This sweep would be particularly useful for measuring extremely short pulses. For example, a spiral beam formed in the X band cavity described would make one complete revolution in about 10^{-10} second and could therefore be used to measure time intervals up to the order of 10^{-11} second.

Experimental X band tubes have been constructed in this Laboratory. The experimental results obtained will be published in a later paper.

APPENDIX A

Solution of the Equations of Motion in a TE_{11n_z} Cavity

The equations of motion (1) can be written using the Laplace transform notation[8] for $0 \leq t \leq t_1$,

$$\left.\begin{aligned} s^2X(s)+\omega_c sY(s) &= -(e/m)\mathfrak{L}(\mathcal{E}_x) \\ s^2Y(s)-\omega_c sX(s) &= -(e/m)\mathfrak{L}(\mathcal{E}_y) \end{aligned}\right\},$$

where s=variable in the complex plane, $\mathfrak{L}$=Laplace transform of, $X(s)$=Laplace transform of $x(t)$, $Y(s)$=Laplace transform of $y(t)$.

The initial conditions assumed are that $x=0$, $\dot{x}=0$, $y=0$, and $\dot{y}=0$ at $t=0$. Then

$$\left.\begin{aligned} X(s) = &-\frac{eE_{0x}}{2m}\frac{\omega_c}{s(s^2+\omega_c^2)} \\ &\times\mathfrak{L}\{\cos[(1+\eta)\omega_0 t+\phi]-\cos[(1-\eta)\omega_0 t+\phi]\} \\ &-\frac{eE_{0y}}{2m}\frac{1}{s^2+\omega_c^2} \\ &\times\mathfrak{L}\{\sin[(1+\eta)\omega_0 t+\phi]-\sin[(1-\eta)\omega_0 t+\phi]\} \end{aligned}\right\}$$

and

$$\left.\begin{aligned} Y(s) = &-\frac{eE_{0x}}{2m}\frac{\omega_c}{s(s^2+\omega_c^2)} \\ &\times\mathfrak{L}\{\sin[(1+\eta)\omega_0 t+\phi]-\sin[(1-\eta)\omega_0 t+\phi]\} \\ &+\frac{eE_{0y}}{2m}\frac{1}{s^2+\omega_c^2} \\ &\times\mathfrak{L}\{\cos[(1+\eta)\omega_0 t+\phi]-\cos[(1-\eta)\omega_0 t+\phi]\} \end{aligned}\right\}.$$

For the linearly polarized mode, $E_{0y}=0$. Then, for $\omega_c \neq (1\pm\eta)\omega_0$,

$$x(t) = -\frac{eE_{0x}}{2m}\mathfrak{L}^{-1}\left[\frac{(1+\eta)\omega_0\cos\phi+s\sin\phi}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} - \frac{(1-\eta)\omega_0\cos\phi+s\sin\phi}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]$$

$$y(t) = -\frac{eE_{0x}}{2m}\omega_c\mathfrak{L}^{-1}\left[\frac{(1+\eta)\omega_0\cos\phi+s\sin\phi}{s(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} - \frac{(1-\eta)\omega_0\cos\phi+s\sin\phi}{s(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]$$

where $\mathfrak{L}^{-1}$ means the inverse Laplace transform. By taking the inverse and letting $\omega_0=\omega_c$, i.e., $\kappa=1$, the Eqs. (3) are obtained.

For the circularly polarized driving field, $E_{0y}\neq 0$ and $\phi=0$, then

$$x(t) = -\frac{eE_{0x}}{2m}\mathfrak{L}^{-1}\left[\frac{(1+\eta)\omega_0}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} - \frac{(1-\eta)\omega_0}{(s^2+\omega_c^2)(s^2-(1-\eta)^2\omega_0^2)}\right] - \frac{eE_{0y}}{2m}\mathfrak{L}^{-1}\left[\frac{\omega_c}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} - \frac{\omega_c}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]$$

$$y(t) = -\frac{eE_{0x}}{2m}\mathfrak{L}^{-1}\left[\frac{\omega_c(1+\eta)\omega_0}{s(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} - \frac{\omega_c(1-\eta)\omega_0}{s(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right] - \frac{eE_{0y}}{2m}\mathfrak{L}^{-1}\left[\frac{s}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)} + \frac{s}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]$$

By taking the inverse of these expressions, and letting $\omega_0=\omega_c$, i.e., $\kappa=1$, and $E_{0x}=E_{0y}$, Eqs. (5) are obtained. Equations (7) are obtained when E_{1x} is substituted for E_{0x} and $-E_{2yx}$ is substituted for E_{0y}.

APPENDIX B

Solution of the Equations of Motion in a TE_{11n_z} Cavity Driven by a Spiraling Electron Beam

The equations of motion (1) can be written in Laplace transform notation[8] for $0 \leq t \leq t_1$,

$$\left.\begin{aligned} s^2X(s)-\dot{x}_0+\omega_c sY(s)-\omega_c y_0 &= -(e/m)\mathfrak{L}(\mathcal{E}_x) \\ s^2Y(s)-y_0 s-\omega_c sX(s) &= -(e/m)\mathfrak{L}(\mathcal{E}_y) \end{aligned}\right\},$$

[7] C. L. Cuccia, "The electron coupler—A developmental tube for amplitude modulation and power control at ultra-high frequencies," RCA Rev. Vol. X, 2, 270–303 (1949).

[8] M. F. Gardner and J. L. Barnes, *Transients in Linear Systems* (John Wiley and Sons, Inc., New York, 1942).

where the initial conditions at $t=0$ are: $x=0$, $y=y_0$, $\dot{x}=\dot{x}_0$, $\dot{y}=0$. Then,

$$x(t)=\mathfrak{L}^{-1}\left\{+\frac{\dot{x}_0}{s^2+\omega_c^2}+\frac{eE_{2x}}{2m}\left[+\frac{(1+\eta)\omega_0}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)}-\frac{(1-\eta)\omega_0}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]\right.$$
$$\left.+\frac{eE_{2y}}{2m}\left[\frac{\omega_c}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)}-\frac{\omega_c}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]\right\}$$

$$y(t)=\mathfrak{L}^{-1}\left\{\frac{y_0 s}{s^2+\omega_c^2}+\frac{\omega_c^2 y_0+\omega_c\dot{x}_0}{s(s^2+\omega_c^2)}-\frac{eE_{2x}}{2m}\left[\frac{\omega_c(1+\eta)\omega_0}{s(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)}-\frac{\omega_c(1-\eta)\omega_0}{s(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]\right.$$
$$\left.+\frac{eE_{2y}}{2m}\left[+\frac{s}{(s^2+\omega_c^2)(s^2+(1+\eta)^2\omega_0^2)}-\frac{s}{(s^2+\omega_c^2)(s^2+(1-\eta)^2\omega_0^2)}\right]\right\}.$$

By taking the inverse and letting $\omega_0=\omega_c$, i.e., $\kappa=1$, Eqs. (11) are obtained.

APPENDIX C

Evaluation of $\frac{1}{2}\epsilon_0\int E^2 d\tau$ in TE_{11n_z} Cavity

The E-field in a TE_{11n_z} cavity is given by[9]

$$\left.\begin{aligned}E_r&=-A_0\frac{1}{k_c r}J_1(k_c r)\sin\theta\sin\frac{2\pi n_z z}{\lambda_g}\\ E_\theta&=A_0 J_1'(k_c r)\cos\theta\sin\frac{2\pi n_z z}{\lambda_g}\end{aligned}\right\},\qquad\text{(C-1)}$$

where n_z is an integer. Then

$$\frac{1}{2}\epsilon_0\int E^2 d\tau=\frac{\epsilon_0 A_0^2}{2}\int_0^{(n_z\lambda_g/2)}\sin^2\frac{2\pi n_z z}{\lambda_g}dz$$
$$\times\int_0^a\int_0^{2\pi}\left[\frac{1}{k_c^2 r^2}J_1^2(k_c r)\sin^2\theta+J_1'^2(k_c r)\cos^2\theta\right]r\,dr\,d\theta$$
$$=\frac{\epsilon_0 A_0^2}{2}\frac{n\lambda_g\pi}{4k_c^2}\left[\int_0^{k_c a}\frac{1}{\rho}J_1^2(\rho)d\rho+\int_0^a\rho J_1'^2(\rho)d\rho\right],\qquad\text{(C-2)}$$

where a is the radius of the cavity, and $\rho=k_c r$. The last integral in (C-2) becomes

$$\int_0^{k_c a}\rho J_1'^2(\rho)d\rho=-\frac{J_1^2(k_c a)}{2}+k_c a J_0(k_c a)J_1(k_c a)$$
$$+\frac{J_0^2(k_c a)}{2}-\frac{1}{2}-\frac{k_c^2 a^2}{2}[J_1^2(k_c a)$$
$$-J_0(k_c a)J_2(k_c a)].\qquad\text{(C-3)}$$

The first integral in (C-2) cannot be integrated in a closed form, but the integrand can be expanded in a series and then integrated. For an upper limit of integration less than 2, only the first six terms are required for three-place accuracy. Thus

$$\int_0^{k_c a}\frac{1}{\rho}J_1^2(\rho)d\rho$$
$$\doteq\left[\frac{\rho^2}{8}-\frac{\rho^4}{64}+\frac{5\rho^6}{4608}-\frac{7\rho^8}{147456}+\frac{\rho^{10}}{737{,}280}\right]_0^{k_c a}.\qquad\text{(C-4)}$$

Then the stored energy integral can be written

$$\tfrac{1}{2}\epsilon_0\int E^2 d\tau=\tfrac{1}{2}\epsilon_0 A_0^2\tau_0'=\tfrac{1}{2}\epsilon_0 E_0^2\tau_0$$

where τ_0' is found from the evaluation of (C-2) using expressions (C-3) and (C-4). Then $\tau_0=4\tau_0'$ because $A_0=2E_0$.

For an X band TE_{111} cavity, where $f_0=9000$ mc, $a=1.75\times10^{-2}$ meter, and $\lambda_g=4\times10^{-2}$ meter. Then $k_c a=1.85$, and the approximation condition is satisfied. For these numerical values, $\tau_0=2.28\times10^{-6}$ meter.[3]

[9] S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio* (John Wiley and Sons, Inc., New York, 1944).

Geometrical Factors Affecting the Contours of X-Ray Spectrometer Maxima. II. Factors Causing Broadening*

LEROY ALEXANDER
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received July 25, 1949)

By applying Stokes' Fourier transform method for the analysis of diffraction maxima it is shown that the pure diffraction contour generated by a crystallite size distribution is apt to be approximated rather closely by the function $1/(1+k^2\phi^2)$. In the case of the x-ray spectrometer this pure diffraction contour is broadened significantly by the action of the following five geometrical factors: (I) the x-ray source width, (II) flat rather than curved sample surface, (III) vertical divergence of the x-ray beam, (IV) penetration of the sample by the beam, and (V) the receiving slit width.

The broadening of the pure diffraction contour due to the action of each of the five factors and the breadth of the final contour generated by the instrument can be deduced by employing the convolution approach suggested by Spencer. The effect of each instrumental factor is expressed by a convolution equation of the form

$$f_i(\phi)=\int_{-\infty}^{+\infty} W_i(\zeta)f_{i-1}(\phi-\zeta)d\zeta,$$

in which ϕ is the angular displacement from twice the ideal Bragg angle, 2θ, f_{i-1} is the contour before the action of the ith geometrical factor, W_i is the form of the ith geometrical factor, and f_i is the contour after the action of W_i on f_{i-1}. Starting with a pure diffraction contour of the form $1/(1+k^2\phi^2)$, generalized broadening curves are derived for the effect of each of the five geometrical factors. Using these curves it is possible to predict the breadth of the final diffraction contour generated by the spectrometer from an initial contour of any breadth.

I. INTRODUCTION

IN addition to its theoretical interest this subject commands attention because of the very general employment of the diffraction contour width as a criterion of crystalline size. The problem is essentially that of deducing the nature of the pure diffraction contour free of instrumental contributions. Several investigators have suggested methods of attacking the problem,[1–6] the most directly applicable one being that of Stokes.[5]

In the present paper each of the geometrical factors causing significant broadening will be discussed separately, and it will be indicated how the net broadening can be estimated as a result of the combined action of the several factors. In the following paper (this issue) these findings are applied to the problem of more accurately determining the crystallite sizes of powders from contour broadening measurements.

When freed of instrumental effects, the breadth of a contour approaches zero for the case of crystallite dimensions greater than 1000A and for moderate Bragg angles (such as must be resorted to in conventional spectrometers). However, at Bragg angles close to 90° the pure contour breadth is measurable for crystallite dimensions as large as 2000A.

For a powder spectrometer in good adjustment the instrumental factors causing appreciable broadening are the following: (I) width of x-ray source (or "source" slit if two slits are used), (II) flat rather than curved sample surface, (III) vertical divergence, (IV) penetration of the sample by the beam, and (V) width of the receiving slit. In the older Norelco model in which Soller slits are not employed, factor (I) is usually dominant while (IV) and (V) may or may not be important. In the newer models vertical divergence is practically eliminated by the use of Soller slits, while at the same time very narrow sources are employed. For this case factor (I) becomes less important and (III) can be neglected.

II. THE MATHEMATICAL EXPRESSION OF INSTRUMENTAL EFFECTS

Spencer[2,3] has pointed out that the effect of an apparatus is to transform a function $f(x)$ into a function $F(x)$ which is directly observable. Assuming that the superposition theorem holds, $F(x)$ is the convolution (*Faltung*) of $f(x)$ and the weight function, $W(x)$, of the apparatus, *viz.*,

$$F(x)=\int_{-\infty}^{+\infty} W(u)f(x-u)du. \qquad (1)$$

This expression has also been derived by Jones[7] for the special case of Debye-Scherrer powder diffraction con-

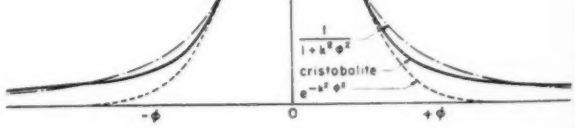


FIG. 1. Comparison of the pure diffraction contour of cristobalite with the functions $1/(1+k^2\phi^2)$ and $\exp(-k^2\phi^2)$.

* Presented in part before the American Society for X-Ray and Electron Diffraction at Cornell University, Ithaca, New York, on June 25, 1949. For Paper No. I, "Factors Causing Asymmetry," see J. App. Phys. **19**, 1068 (1948).

[1] Roy C. Spencer, Phys. Rev. **38**, 618 (1931).
[2] Roy C. Spencer, Phys. Rev. **55**, 239 (1939).
[3] Roy C. Spencer, J. App. Phys. **20**, 413 (1949).
[4] C. G. Shull, Phys. Rev. **70**, 679 (1946).
[5] A. R. Stokes, Proc. Phys. Soc. **61**, 382 (1948).
[6] A. Kochendorfer, Zeits. f. Krist. **105**, 393 (1944).
[7] F. W. Jones, Proc. Roy. Soc. **A166**, 16 (1938).

tours. The auxiliary variable u has the same dimensions as x and its introduction in Eq. (1) is a characteristic feature of the superposition theorem. The weight function, $W(x)$, is to be regarded as a function which expresses the sum total of the apparatus effects upon the pure function which is being measured. Alternately, the instrumental weight function may be split up into each of the several instrumental factors, which it may be more convenient to analyze separately. The ultimate response curve of the apparatus may then be obtained by the successive operation of each specific weight function upon the response curve generated by the action of the preceding specific function. Thus, for n instrumental factors and weight functions W_I, W_{II}, $W_{III}, \cdots W_n$, we would have the generation of the respective contours $f_1, f_2, f_3, \cdots, f_n = F$, according to the successive action of these weight functions:

$$\left.\begin{aligned} f_1(x) &= \int_{-\infty}^{+\infty} W_I(u) f(x-u)\,du, \\ f_2(x) &= \int_{-\infty}^{+\infty} W_{II}(u) f_1(x-u)\,du, \\ f_3(x) &= \int_{-\infty}^{+\infty} W_{III}(u) f_2(x-u)\,du, \\ f_n(x) = F(x) &= \int_{-\infty}^{+\infty} W_n(u) f_{n-1}(x-u)\,du. \end{aligned}\right\} \quad (2)$$

It should be borne in mind that in Eqs. (1) and (2), $f(x-u)$ is the pure function which we are seeking to measure and which was originally written simply as $f(x)$. It must be kept in mind that the symbol f denotes a specific function, regardless of the notation used for the independent variable (x or $x-u$).

The elucidation of contour broadening in an x-ray powder spectrometer, then, resolves itself into the twofold task of (a) deducing or assuming satisfactorily the nature of the pure diffraction contour, $f(x)$, and (b) determining the proper forms of the five instrumental weight functions W_I, W_{II}, W_{III}, W_{IV}, and W_V, corresponding to the five factors listed above which contribute to contour breadth. Once this has been done, Eqs. (2) can be employed to synthesize the final contour generated by the instrument starting from any original pure diffraction contour.

III. THE PURE DIFFRACTION CONTOUR

Case I. Crystallite Size > 1000A

For moderate Bragg angles the pure contour width is imperceptible and may be neglected. Hence, in Eq. (2) $f(x-u)$ is effectively a constant and $f_1(x) = W_I(x)$, so that for this case the first convolution equation drops out and we start with the action of W_{II}, the flat sample factor, on W_I, the x-ray source contour.

TABLE I. Pure diffraction contours of miscellaneous materials showing crystallite size broadening. The first five were analyzed by the Fourier method of Stokes. Degree of resemblance to $e^{-k^2\phi^2}$ or $1/(1+k^2\phi^2)$ is indicated as an approximate percentage.

No.	Material	Origin	Percentage resemblance to $e^{-k^2\phi^2}$	Percentage resemblance to $1/(1+k^2\phi^2)$
1	MgO	Decomposition of magnesium carbonate at 600°C.	20	80
2	CaF_2	Presumably precipitated from aqueous solution.	50	50
3	Cristobalite	Thermal treatment of diatomaceous earth (amorphous SiO_2).	20	80
4	Basic calcium phosphate	Precipitated from aqueous solution.	10	90
*5	Cold worked copper	Specific information not published.	20	80
**6	Nickel catalyst	Alkaline treatment of Raney alloy.	10	90

* This contour published by Stokes.
** This contour not analyzed by the Fourier method because its width was so great as to render the instrumental contributions negligible.

Case II. Crystallite Size < 1000A

The pure diffraction contour is a function of the mean crystallite shape[8] and of the size distribution. Obviously these factors will vary widely with the chemical constitution and origin of any particular powder sample. The investigations of Jones[7] tend to show that for a crystallite size distribution, the diffraction contour is given more closely by the function $1/(1+k^2x^2)$ than by a normal distribution curve ($e^{-k^2x^2}$). Stokes[5] has recently described a Fourier transform method for deducing the nature of the pure diffraction contour free of instrumental contributions. This method makes it possible for the first time to replace the somewhat speculative notions of the past with dependable information about the nature of the pure diffraction contour.

Using a Patterson-Tunell[9] set of strips and stencils for the rapid summation of Fourier series, the pure diffraction contours of a number of materials have been determined by the procedure of Stokes with the results summarized in Table I. In Fig. 1 the rather typical case of the cristobalite contour is portrayed. Its comparatively close coincidence with the function $1/(1+k^2\phi^2)$ is evident, and it will be noted from Table I that five of the six materials showed equally good or even better agreement with this function. Only in the case of CaF_2 was the agreement poorer, and even then the pure diffraction contour did not resemble $e^{-k^2\phi^2}$ more closely than $1/(1+k^2\phi^2)$. The most fundamental point of difference between these two mathematical functions is the much more gradual decrease in magnitude of the latter as ϕ becomes large. As suggested by Jones,[7] it is not surprising that this function should more accurately fit the case of the diffraction contour of a crystallite size distribution. There are theoretical reasons for $e^{-k^2\phi^2}$ being the better choice if all the

[8] A. R. Stokes and A. J. C. Wilson, Proc. Camb. Phil. Soc. **38**, 313 (1942).
[9] A. L. Patterson and G. Tunell, Am. Mineralogist **27**, 655 (1942).

crystallites are of the same size, whereas for a distribution of sizes the large crystallites will tend to sharpen the peak while the small ones will broaden the base of the contour.

In view of the above considerations together with the experimental data of Table I, there appears to be good justification for choosing the simple mathematical function $1/(1+k^2\phi^2)$ to approximate the pure diffraction contour generated by a typical crystallite size distribution.

Having thus decided on a suitable form for the pure contour function, $f(x-u)$, of Eqs. (2), we now proceed to evaluate each of the five weight functions corresponding to the five geometrical factors which act to modify it. In order to retain maximum generality in the following analysis, we shall discuss Case II, in which the pure diffraction contour possesses appreciable breadth. In the ensuing treatment as well as in the preceding three paragraphs the generalized parameter x of Eqs. (1) and (2) is replaced by the appropriate quantity ϕ, which is the angular displacement from the theoretical angle, 2θ, of a given reflection. Unless otherwise indicated the term *width*, or *breadth*, of a contour will refer to the width at half maximum intensity.

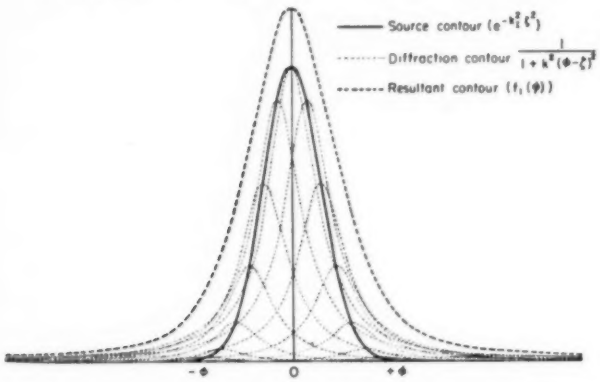


FIG. 2. Schematic representation of the convolution of the source contour and a pure diffraction contour:

$$f_1(\phi)=\int_{-\infty}^{+\infty}\frac{\exp(-k_I^2\zeta^2)d\zeta}{1+k^2(\phi-\zeta)^2}.$$

The case pictured is that of $w/w_I=0.85$ and $w_1/w=1.80$.

IV. EFFECT OF THE X-RAY SOURCE WIDTH (W_I)

In the past it has been customary to assume that the distribution of energy across the focus of the x-ray tube is either uniform or of the form $e^{-k^2\phi^2}$.[10] Pinhole photographs of tube foci usually reveal a more or less irregular intensity distribution which does not conform to either of these assumptions, although it commonly approaches the exponential state more or less closely. With a focusing powder spectrometer possessing a relatively wide source (of the order of $0.20°2\theta$), it is possible to observe the effective source contour with little distortion by scanning a diffraction contour produced by some material of large crystallite size in order to eliminate the pure contour breadth. At the same time it is necessary to keep the x-ray beam divergence small and to employ a narrow receiving slit. Under these conditions the factors other than source width become of minor importance in influencing the observed contour. Using a Norelco Geiger-Counter Spectrometer Type No. 12021, the copper and iron anode tubes used in the present investigation were thus shown to possess source configurations approaching the normal distribution curve, $e^{-k^2\phi^2}$, very closely. Such an intensity distribution will be assumed in the present paper, although it must be remembered that a radical departure from this type of configuration, which might well be found for some particular x-ray tubes, would lead to a corresponding modification of the analysis to follow.

We may write for the weight function of the source contour

$$W_I(\phi)=\exp(-k_I^2\phi^2), \tag{3}$$

in which the value of k_I determines the width of the contour and from which we have eliminated any amplitude coefficient since it has no bearing on the contour shape. The pure contour is assumed to be of the form

$$f(\phi)=1/(1+k^2\phi^2), \tag{4}$$

so that we obtain for the convolution of W_I and f

$$_1(\phi)=\int_{\infty}^{+\infty}W_I(\zeta)f(\phi-\zeta)d\zeta$$

$$=\int_{-\infty}^{+\infty}\frac{\exp(-k_I^2\zeta^2)d\zeta}{1+k^2(\phi-\zeta)^2}. \tag{5}$$

In terms of the widths of the W_I and f contours at half maximum intensity, w_I and w, we find for k_I and k respectively:

$$(W_I)_{\frac{1}{2}}=(W_I)_0/2=\tfrac{1}{2}=\exp(-k_I^2\phi_{\frac{1}{2}}^2)$$

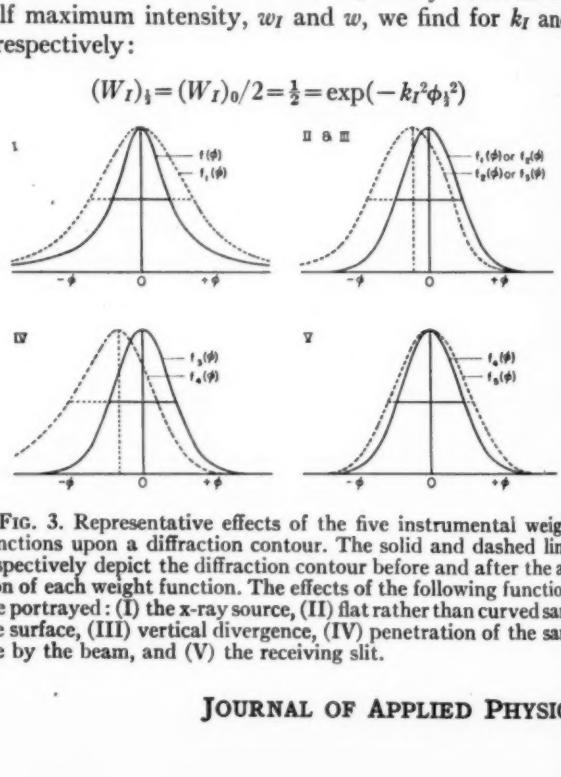


FIG. 3. Representative effects of the five instrumental weight functions upon a diffraction contour. The solid and dashed lines respectively depict the diffraction contour before and after the action of each weight function. The effects of the following functions are portrayed: (I) the x-ray source, (II) flat rather than curved sample surface, (III) vertical divergence, (IV) penetration of the sample by the beam, and (V) the receiving slit.

[10] A. Taylor and H. Sinclair, Proc. Phys. Soc. **57**, 108 (1945).

or

$$k_I=(\log_e 2)^{\frac{1}{2}}/\phi_{\frac{1}{2}}=2(\log_e 2)^{\frac{1}{2}}/w_I=1.67/w_I, \quad (6)$$

and

$$f_{\frac{1}{2}}=f_0/2=\tfrac{1}{2}=1/(1+k^2\phi_{\frac{1}{2}}^2)$$

or

$$k=1/\phi_{\frac{1}{2}}=2/w. \quad (7)$$

Using the procedure described above for determining the approximate form of the source intensity distribution, it was found that w_I for the copper anode tube was about 0.20° (in $\Delta 2\theta$, or ϕ, units). This agreed substantially with a geometrical deduction of the angular source width from the manufacturer's specifications for the size of the focus. In the improved spectrometers with Soller slits and vertical focal spots, W_I is ordinarily very much smaller.

An attempt has been made in Fig. 2 to picture the meaning of the convolution equation, (5). The source contour acts as the envelope for an infinite number of diffraction contours (represented by a finite number in Fig. 2) all of the same shape but differing in amplitude. The resultant intensity at any point ϕ is the sum of the intensities of the various diffraction contours at this point. In Fig. 2 the amplitude of the resultant contour has been greatly scaled down to permit plotting it in the same diagram with the source and diffraction contours. It will be observed that the resultant contour is much broader than the original diffraction contour.

Equation (5) cannot be integrated readily, but sufficiently accurate solutions can be obtained by evaluating the equivalent summation at small intervals:

$$f_1(\phi)=\sum_{-\infty}^{+\infty}\frac{\exp(-k_I^2\zeta^2)}{1+k^2(\phi-\zeta)^2}\Delta\zeta. \quad (8)$$

The calculations are facilitated by graphing the denominator function for values of $(\phi-\zeta)$ and by utilizing standard tables of the normal distribution function, $(2\pi)^{-\frac{1}{2}}e^{-(\frac{1}{2})x^2}$, for obtaining the numerator. This is accomplished by converting the exponential function from $\exp(-k_I^2\zeta^2)$ to $\exp[-\frac{1}{2}(2^{\frac{1}{2}}k_I\zeta)^2]$, permitting the value of the numerator to be obtained as a function of the argument $2^{\frac{1}{2}}k_I\zeta$. When $w_I=0.20°$ the interval $\Delta\zeta$ should be about 0.01°. Figure 3(I) depicts the action of a source of breadth $w_I=0.20°$ in broadening a pure diffraction contour of breadth $w=0.20°$ to a resultant contour for which $w_1=0.327°$.

The relative broadening of the pure contour is a function only of the ratio of the breadth of the pure contour to the breadth of the x-ray source contour, w/w_I. The relative broadening is the ratio of the breadth of the contour after the action of the source factor, which we may denote by w_1, to its original breadth, w. Figure 4(I) shows this broadening ratio, w_1/w, as a function of w/w_I, the location of the curve having been precisely determined by the application of Eq. (8) to a number of pairs of values of w and w_I [and, therefore, also of k and k_I by Eqs. (6) and (7)].

V. EFFECT OF FLAT RATHER THAN CURVED SAMPLE SURFACE (W_{II})

From Fig. 5 it is seen that an x-ray beam of horizontal divergence angle α and impinging upon a flat sample surface at the angle θ (as in conventional spectrometers) will irradiate a length of sample which is approximately given by

$$2s=\alpha d/\sin\theta, \quad (9)$$

where d is the distance from source to sample. This formula is very accurate for the usual small values of α of not more than 2° or 3°. In actual practice the effective beam divergence at low Bragg angles is limited by the length of the sample, l, so that for $\sin\theta<\alpha d/l$ the effective divergence is $\alpha=(l/d)\sin\theta$. To simplify the present discussion, however, the sample surface will be assumed to be very large so that (9) holds at all values of θ. We shall first investigate the effect of sample flatness in broadening the focus of the diffracted rays at the receiving slit.

Figure 6 shows how this happens. The flat sample, S_1-S_2, is tangent to the focusing circle at its midpoint, S_0. A central ray FS_0, which impinges at the Bragg angle θ upon the surface, is diffracted (also at the Bragg angle θ) to the theoretical focus at G. But a non-central ray, FS', impinging at an angle $\theta+\Delta\theta$, is diffracted at

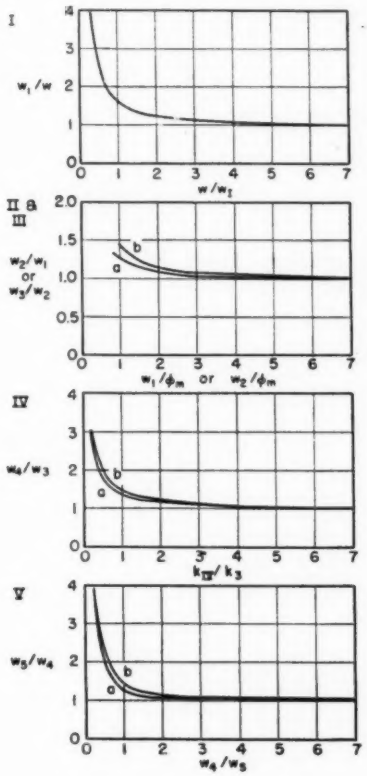


FIG. 4. Broadening ratios for the five instrumental weight functions, including (I) the x-ray source, (II) flat rather than curved sample surface, (III) vertical divergence, (IV) penetration of the sample by the beam, and (V) the receiving slit. Curves (a) and (b) apply respectively to $\exp(-k^2\phi^2)$ and $1/(1+k^2\phi^2)$ contours.

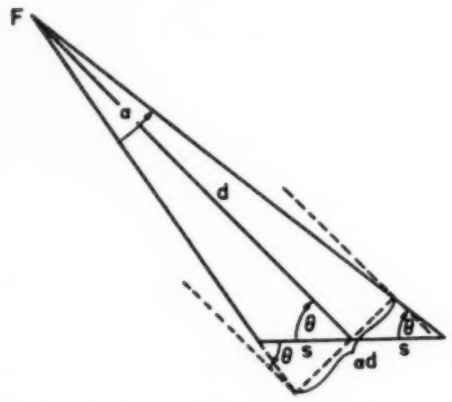


FIG. 5. Irradiation of a flat sample surface by a beam of divergence angle α.

the angle $\theta-\Delta\theta$, and arrives at some point G_1 which is displaced from the ideal focus by the linear interval GG_1. This displacement is proportional to z, the displacement of the point S' on the flat sample surface from the focusing circle. We now employ the approximate formula which expresses the length of the sagitta in terms of the circle radius r and the ordinate h (see Fig. 6, insert)

$$x=h^2/2r, \tag{10}$$

which is rather accurate for $h\ll r$. Applying this relationship to the situation of Fig. 6, we find that

$$z=h^2/2R, \tag{11}$$

where h is the distance S_0S' and R is the radius of the focusing circle. Since $R=d/2\sin\theta$,[11] this gives for z

$$z=h^2\sin\theta/d. \tag{12}$$

The displacement GG_1 from the ideal focus is then

$$GG_1=2z\cos(\theta+\Delta\theta)=(h^2/d)\sin2\theta, \tag{13}$$

or in terms of the angular deviation from the theoretical diffraction angle 2θ

$$\begin{aligned}\phi_{GG_1}&=-(h^2/d^2)\sin2\theta \text{ radians}\\&=-57.3(h^2/d^2)\sin2\theta \text{ degrees.}\end{aligned} \tag{14}$$

The minus sign indicates that the displacement ϕ is in the -2θ direction.

The limiting displacement of the diffracted rays from the ideal focus G corresponds to rays diffracted from the points S_1 and S_2, for which $h=s$. Then

$$\phi_m=-(s^2/d^2)\sin2\theta \text{ radians},$$

or, since $s\cong\alpha d/2\sin\theta$,

$$\begin{aligned}\phi_m&=-\frac{\alpha^2}{4\sin^2\theta}\cdot 2\sin\theta\cos\theta\\&=-\tfrac{1}{2}\alpha^2\cot\theta \text{ radians.}\end{aligned} \tag{15}$$

[11] L. Alexander, J. App. Phys. **19**, 1068 (1948).

For ϕ and α in degrees this becomes

$$\phi_m=-\frac{\alpha^2\cot\theta}{114.6} \text{ degrees.} \tag{16}$$

This equation shows that the sharpness of focus (as affected by the flat sample factor) improves rapidly with decreasing beam divergence and increasing diffraction angle. If the finite length, l, of the sample surface is also taken into account, at angles for which $\sin\theta<\alpha d/l$ the limiting displacement of the diffracted rays from the ideal focus G is

$$\phi_m=-\tfrac{1}{2}(l^2/114.6d^2)\sin2\theta \text{ degrees.} \tag{17}$$

In practice, then, the sharpness of focus as a result of the flat sample effect is poorest (ϕ_m is a maximum) when $\sin\theta=\alpha d/l$ and it improves at both smaller and larger angles. We now need to determine the flat sample weight function, $W_{II}(\phi)$. From Eq. (13) or from an inspection of Fig. 6 it is seen that a displacement GG_1 (or ϕ_{GG_1}) is proportional to the sagitta variable z. But the intensity at a displacement ϕ from the theoretical focus is proportional to the energy diffracted in the interval Δz_ϕ (or $\Delta\phi_\phi$). This energy is proportional to the amount of sample surface diffracting into the increment Δz (or $\Delta\phi$), which is Δh_z (or Δh_ϕ). It follows, therefore, that the $W_{II}(\phi)$ function is proportional to $(\Delta h/\Delta z)$. But it has already been shown that we make negligible error in using the sagitta relationship for z [Eq. (11)], so that in the limit

$$W_{II}(\phi)\propto dh/dz=R/h=(R/2z)^{\frac{1}{2}}.$$

Since we are not interested in the absolute magnitude of the amplitude but only in the nature of the contour, and remembering that ϕ is proportional to z, we may dispense with a proportionality constant and write

$$W_{II}(\phi)=\phi^{-\frac{1}{2}}. \tag{18}$$

This function has the objection of an infinite ordinate at $\phi=0$, but the difficulty can be circumvented by evaluating the resulting convolution integral as a summation and omitting the zero term. For the convolution of $f_1(\phi)$ and $W_{II}(\phi)$ we then have

$$\begin{aligned}f_2(\phi)&=\int_{-\infty}^{+\infty}W_{II}(\zeta)f_1(\phi-\zeta)d\zeta\\&=\int_{\phi_m}^{0}|\zeta|^{-\frac{1}{2}}\exp[-k_1^2(\phi-\zeta)^2]d\zeta,\end{aligned} \tag{19}$$

assuming the f_1 contour to approximate an error curve. To evaluate $f_2(\phi)$ we go to the equivalent summation

$$f_2(\phi)=\sum_{\zeta=\phi_m}^{\zeta=\phi_1}|\zeta|^{-\frac{1}{2}}\exp[-k_1^2(\phi-\zeta)^2]\Delta\zeta \tag{20}$$

and sum at a relatively large number of ζ values (small

$\Delta\zeta$), omitting the $\zeta=0$ term. This omission has the effect of shifting the center of gravity of the contour slightly, but it will furnish a rather accurate contour shape provided $\Delta\zeta$ is kept small, e.g., of the order of $\phi_m/40$ or less. It will be noticed that ζ has been assigned to a range of negative rather than positive values. This is consistent with the sign convention of Eqs. (15) and (16) and is correct because the function W_{II} distorts a diffraction contour in the -2θ direction.

As in Part IV, the computations for actual experimental conditions are facilitated by graphing one function, this time $W_{II}(\zeta)=|\zeta|^{-\frac{1}{2}}$, and utilizing standard tables of the normal distribution function to evaluate $\exp[-k_1^2(\phi-\zeta)^2]$, after first writing the exponent as $-\frac{1}{2}[2^{\frac{1}{2}}k_1(\phi-\zeta)]^2$. For illustration we may outline briefly the computations for a rather extreme case of broadening, which portrays the nature of the results to be expected in a striking manner. The case is one of large beam divergence ($\alpha=1.41°$) and low Bragg angle ($\theta=5°$). From Eq. (16) we find for the greatest displacement of the diffracted rays from the theoretical focus:

$$\phi_m=-(\alpha^2\cot\theta)/114.6=-2\times11.43/114.6=-0.200°.$$

Assuming that the contour of the x-ray source is an error curve in form and that the width at half maximum intensity is 0.200° (there being no pure diffraction broadening assumed in this case), we find for k_1 by applying Eq. (6):

$$k_1=1.67/w_1=1.67/w_I=1.67/0.200=8.35 \text{ degrees}^{-1}.$$

When summed at intervals of $\Delta\zeta=0.005°$, Eq. (20) becomes

$$f_2(\phi)=\sum_{\zeta=-0.200°}^{\zeta=-0.005°}|\zeta|^{-\frac{1}{2}}\exp[-8.35^2(\phi-\zeta)^2]\times0.005°.$$

Dropping the constant factor 0.005° since it does not affect the nature of the contour, and converting the exponent into the standard error curve form, we have

$$f_2(\phi)=\sum_{\zeta=-0.200°}^{\zeta=-0.005°}|\zeta|^{-\frac{1}{2}}\exp[-\tfrac{1}{2}\{11.8(\phi-\zeta)\}^2]. \quad (21)$$

Figure 3(II) shows the $f_2(\phi)$ contour as computed with Eq. (21) and compares it with the original contour $f_1(\phi)=\exp(-8.35^2\phi^2)$. The peak has been broadened from $w_1=0.200°$ to $w_2=0.249°$, which means a broadening ratio of $w_2/w_1=1.245$. In addition the resulting peak is seen to be decidedly asymmetric and the center of gravity and maximum point have been shifted in the -2θ direction by appreciable amounts.

A consideration of Eq. (21) shows that the broadening ratio w_2/w_1 is a function only of the relative sizes of w_1 and ϕ_m. Accordingly if the $f_2(\phi)$ function is computed for a number of values of the ratio w_1/ϕ_m which cover all anticipated experimental situations, another curve of perfectly general applicability is obtained which is analogous to the curve of Fig. 4(I) developed in Part IV. This curve of w_2/w_1 as a function of w_1/ϕ_m is shown in Fig. 4(II,a). When the function $f_1(\phi-\zeta)$ of Eq. (19) has the form $1/[1+k_1^2(\phi-\zeta)^2]$, curve (b) is obtained. Curve (b) is more appropriate than (a) for cases in which the pure diffraction contour is much wider than the x-ray source contour. When w and w_I do not differ very greatly the broadening will be intermediate between the values predicted by curves (a) and (b).

For moderate values of the beam divergence, and Bragg angles of 10° or more, it is found that the broadening ratio seldom exceeds 1.02. This permits us to conclude that for ordinary experimental conditions the effect of flat sample surface is very minor.

VI. EFFECT OF VERTICAL DIVERGENCE OF THE X-RAY BEAM (W_{III})

In a preceding paper[11] the effect of vertical divergence and other third-dimensional properties of the spectrometer in producing asymmetry was discussed in detail. Referring to Fig. 2 of this reference, in the older spectrometers without Soller slit collimators the usual working value of the x-ray slit height, h_x, is considerably larger than the source height, h_f. Under these conditions we make but little error in assuming that the x-ray source has negligible height and that the centers of the diffraction halos in the plane $G'G$ (refer to Figs. 1 and 3 of the above reference) are uniformly distributed over a vertical interval

$$L=(h_x d/b)(1+\cos2\theta),$$

b and d being respectively the distances from the x-ray source to the x-ray slit and sample. Defining the vertical divergence angle β by h_x/b, this becomes

$$L=\beta d(1+\cos2\theta), \quad (22)$$

β being expressed in radians.

As a first approximation we may assume negligible receiving slit height. In that case, referring to Fig. 3 of the above reference, the maximum displacement of a diffracted ray from the ideal receiver focus is given by the sagitta of the arc of height $\frac{1}{2}L$ and radius r. Em-

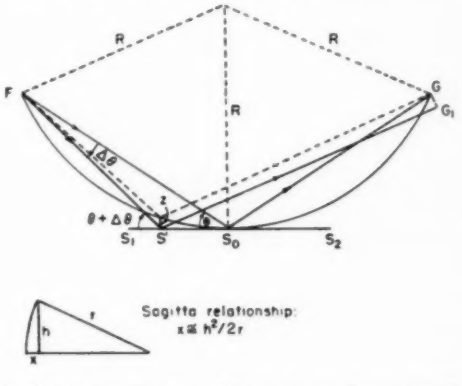


FIG. 6. Displacement of a diffracted ray from the theoretical focus G because of the use of a flat rather than curved sample surface.

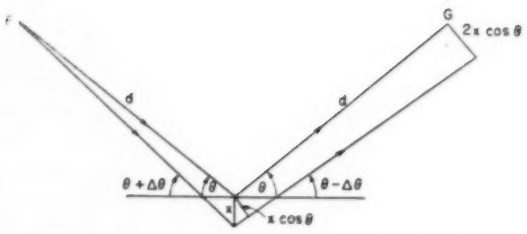


FIG. 7. Displacement of a diffracted ray from the theoretical focus G due to penetration of the sample to a depth x.

ploying the approximate sagitta formula as was done in Part V, we find for this displacement

$$x\ (\text{cm}) = (L/2)^2/2r = L^2/8r; \tag{23}$$

but from Fig. 1 of reference 11 it can be seen that a displacement x in the direction GG' equals an angular displacement ϕ tangent to the spectrometer arc of

$$\phi = x\ (\text{cm}) \times (\cos 2\theta)/d \text{ radians}.$$

Combining this with Eqs. (22) and (23) and noting that $r = d \sin 2\theta$, we obtain

$$\phi_m = -(\beta^2/8)(1+\cos 2\theta)^2 \cot 2\theta \text{ radians}, \tag{24}$$

the negative sign indicating that the displacement from the theoretical angle 2θ is in the negative direction. For ϕ_m and β in degrees Eq. (24) becomes

$$\phi_m = -(\beta^2/8 \times 57.3)(1+\cos 2\theta)^2 \cot 2\theta \text{ degrees}. \tag{25}$$

The preceding analysis demonstrates that, as in Part V, we are again dealing with an instrumental weight function, this time $W_{III}(\phi)$, which varies as the inverse square root of the displacement, ϕ, from the ideal angle of deviation, 2θ. This follows in a manner parallel to that involved in the development of Eq. (18) since the intensity corresponding to a displacement ϕ from the theoretical focus is proportional to the energy diffracted in the interval Δx_ϕ [or $(\Delta\phi)_\phi$], and this energy is proportional to $(\Delta L)_\phi$. Thereupon in the limit

$$W_{III}(\phi) \propto \frac{dL}{dx} = \left(\frac{2r}{x}\right)^{\frac{1}{2}},$$

and, neglecting the constant factor as in Part V, and remembering that x is proportional to ϕ, we again arrive at the weight function,

$$W_{III}(\phi) = \phi^{-\frac{1}{2}}. \tag{26}$$

Assuming $f_2(\phi)$ to approximate an error curve, we then find for the convolution, $f_3(\phi)$, of $W_{III}(\phi)$ and $f_2(\phi)$

$$f_3(\phi) = \int_{-\infty}^{+\infty} W_{III}(\zeta) f_2(\phi - \zeta) d\zeta,$$

or

$$f_3(\phi) = \sum_{\zeta=\phi_m}^{\zeta=\phi_1} |\zeta|^{-\frac{1}{2}} \exp[-k_2^2(\phi-\zeta)^2]\Delta\zeta, \tag{27}$$

which is identical in form to Eq. (20).

It is evident that the remarks made in connection with Eq. (20) apply here, and also that curves (a) and (b) of Fig. 4 (II and III) apply directly in the present case for w_3/w_2 as a function of w_2/ϕ_m. The only difference is that Eq. (25) instead of Eq. (16) is to be used to determine ϕ_m for the particular experimental conditions under consideration.

For Bragg angles of 10° or more and for vertical divergence angles less than 2.4° it is found that the broadening ratio is seldom greater than 1.10 and more commonly it is of the order of 1.05. Evidently vertical divergence is much more important than flat sample surface in producing contour broadening.

VII. EFFECT OF PENETRATION OF THE SAMPLE BY THE X-RAY BEAM (W_{IV})

Consider a ray incident upon the sample at the angle $\theta+\Delta\theta$ and diffracted from a crystallite at a depth x (Fig. 7). If $\Delta\theta$ is neglected because it is very small, the intensity of the emergent diffracted ray, I, relative to that of a ray diffracted under conditions of non-absorption, I_0, is

$$I/I_0 = \exp(-2\mu x/\sin\theta), \tag{28}$$

μ being the linear absorption coefficient of the powder. The linear displacement of this diffracted ray from the theoretical focus at the receiver is evidently $2x\cos\theta$, and the corresponding angular displacement, ϕ_x, is

$$\phi_x = -(114.6x\ \cos\theta)/d \text{ degrees}, \tag{29}$$

the negative sign denoting a displacement in the -2θ direction. Equation (28), then, is an expression of the instrumental weight function, W_{IV}, due to penetration of a thick powder sample by the x-ray beam, and when this is combined with Eq. (29) we obtain W_{IV} expressed as a function of ϕ:

$$W_{IV}(\phi) = e^{k_{IV}\phi}, \tag{30}$$

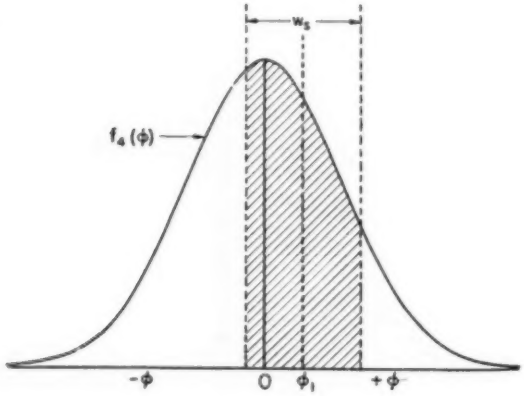


FIG. 8. Representation of a receiving slit of width w_s scanning a diffraction contour $f_4(\phi)$. The resulting intensity is proportional to the area of the contour intercepted (shaded zone).

in which

$$k_{IV}=4\mu d/114.6\sin 2\theta, \tag{31}$$

and ϕ varies between the limits 0 and $-\infty$.

Assuming first that $f_3(\phi)$ approximates an error curve (i.e., that the pure diffraction broadening is slight and that the preceding convolution sequence has not greatly modified the original error curve contour of the x-ray source), we obtain for the convolution, $f_4(\phi)$, of $W_{IV}(\phi)$ and $f_3(\phi)$

$$f_4(\phi)=\int_{-\infty}^{+\infty} W_{IV}(\zeta)f_3(\phi-\zeta)d\zeta,$$

or

$$f_4(\phi)=\int_0^{-\infty} e^{k_{IV}\zeta}\exp[-k_3^2(\phi-\zeta)^2]d\zeta. \tag{32}$$

By combining the two exponential factors and removing a constant factor from the integral, Eq. (32) can be put in the form

$$f_4(\phi)=\exp\left[k_{IV}\left(\phi+\frac{k_{IV}}{4k_3^2}\right)\right]\times\int_0^{-\infty}\exp\left[-\left(k_3\zeta-\frac{k_{IV}+2k_3^2\phi}{2k_3}\right)^2\right]d\zeta.$$

A change of variable from ζ to

$$u=2^{\frac{1}{2}}\left(k_3\zeta-\frac{k_{IV}+2k_3^2\phi}{2k_3}\right)$$

then leads to the result

$$f_4(\phi)=\exp\left[\left(k_{IV}\phi+\frac{k_{IV}^2}{4k_3^2}\right)\Big/2^{\frac{1}{2}}k_3\right]\times\int_l^{-\infty}\exp[-(u^2/2)]du, \tag{33}$$

in which the lower limit of the integral, l, is given by

$$l=(k_{IV}/2^{\frac{1}{2}}k_3)+2^{\frac{1}{2}}k_3\phi. \tag{34}$$

The expression for the convolution $f_4(\phi)$ has thus been reduced to the state of an integral (Eq. (33)) which can be evaluated directly by reference to tables of the normal curve of error. For any particular sample material of linear absorption coefficient μ and for a given spectrometer setting 2θ the value of k_{IV} can be calculated with Eq. (31), while k_3 depends upon the width at half maximum intensity of the error curve contour $f_3(\phi)$ according to a relationship identical with Eq. (6), or

$$k_3=1.67/w_3. \tag{35}$$

Figure 3(IV) depicts the effect of sample penetration upon a contour of the form $f_3(\phi)=\exp[-(k_3^2\phi^2)]$ for the particular case of $k_{IV}/k_3=1$. This case was chosen because it is typical of experimental conditions under which the effects of sample penetration are conspicuous. The given values of k are satisfied by either of the following sets of experimental conditions, the first corresponding to the old Norelco spectrometer with wide source ($w_I=0.20°$) and the second to a spectrometer with a relatively narrow source ($w_I=0.07°$):

I. Width of source contour: $w_I=0.20°$
Bragg angle : $\theta=15°$
Linear absorption coefficient of powder : $\mu=9$ or $\begin{cases}\theta\to 45°\\ \mu=18\end{cases}$

(CuKα radiation and powders consisting of metal-organic compounds or mixtures of organic compounds and common siliceous minerals)

II. Width of source contour: $w_I=0.07°$
Bragg angle : $\theta=15°$
Linear absorption coefficient of powder : $\mu=27$ or $\begin{cases}\theta\to 45°\\ \mu=54\end{cases}$

(CuKα radiation and powders prepared from common siliceous minerals)

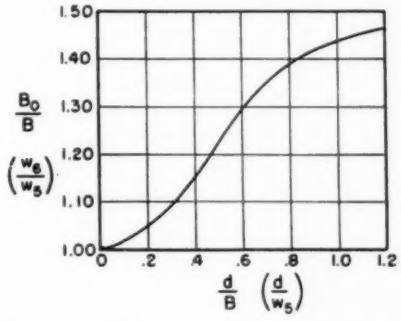


FIG. 9. Inverse of Jones' curve for correcting an observed contour breadth for Kα doublet broadening. B_0=observed breadth, B=breadth corrected for Kα doublet, d=angular separation of α_1 and α_2.

The broadening ratio for this case, w_4/w_3, is a function only of the ratio of the contour constants, k_{IV}/k_3. Curve (a) of Fig. 4(IV) depicts this function for the case just discussed wherein the $f_3(\phi)$ contour is given by $\exp[-(k_3^2\phi^2)]$. When $f_3(\phi)$ is given by $1/(1+k_3^2\phi^2)$, a condition approached by materials giving considerable pure diffraction broadening, the convolution integral cannot be evaluated directly as in the case of Eq. (32) but must be treated as the equivalent summation in the manner described in earlier sections. In this case curve (b) of Fig. 4(IV) is obtained.

If the absorption coefficient of the sample is low and if θ is not very small, sample penetration can evidently be an important source of contour broadening. Thus, in the example cited above for a powder of large crystallite size the broadening ratio was found to be 1.40. In the newer spectrometers with narrow sources this instrumental factor is relatively much more important than in the old Norelco instrument. For example, if $k_{IV}/k_3=2.86$ for $w_I=0.20°$, it will be decreased to 1.00 for $w_I=0.07°$, and reference to curve (a) of Fig. 4(IV) shows that this increases the broadening ratio from 1.14

TABLE II. Summarizing table of the effects of the instrumental weight functions upon the diffraction contour.*

Instrumental weight function		Contour function		Contour width		Form of convolution expression	Constants		Integral limits
		Initial	Final	Initial	Final	General / Spectrometer case	Instrumental	Diffraction contour	
W_I	Source	$f(\phi)$	$f_1(\phi)$	w	w_1	$f_1(\phi)=\int_{-\infty}^{+\infty} W_I(\zeta)f(\phi-\zeta)d\zeta$ $\int_{-\infty}^{+\infty}\frac{\exp(-k_I^2\zeta^2)d\zeta}{1+k^2(\phi-\zeta)^2}$	$k_I=\frac{1.67}{w_I}$	$k=\frac{2}{w}$	
W_{II}	Flat sample	$f_1(\phi)$	$f_2(\phi)$	w_1	w_2	$f_2(\phi)=\int_{-\infty}^{+\infty} W_{II}(\zeta)f_1(\phi-\zeta)d\zeta$ (A) $\int_{\phi_m}^{\phi}\lvert\zeta\rvert^{-4}\exp[-k_1^2(\phi-\zeta)^2]d\zeta$		$k_1=\frac{1.67}{w_1}$ $k_1=\frac{2}{w_1}$	$\phi_m=-\frac{\alpha^2\cot\theta}{114.6}$
W_{III}	Vertical divergence	$f_2(\phi)$	$f_3(\phi)$	w_2	w_3	$f_3(\phi)=\int_{-\infty}^{+\infty} W_{III}(\zeta)f_2(\phi-\zeta)d\zeta$ (B) $\int_{\phi_m}^{\phi}\frac{\lvert\zeta\rvert^{-4}d\zeta}{1+k_2^2(\phi-\zeta)^2}$		$k_2=\frac{1.67}{w_2}$ $k_2=\frac{2}{w_2}$	
W_{IV}	Sample penetration	$f_3(\phi)$	$f_4(\phi)$	w_3	w_4	$f_4(\phi)=\int_{-\infty}^{+\infty} W_{IV}(\zeta)f_3(\phi-\zeta)d\zeta$ (A) $\int_{-\infty}^{\phi} e^{k_{IV}\zeta}\exp[-k_3^2(\phi-\zeta)^2]d\zeta$	$k_{IV}=\frac{4\mu d}{114.6\sin 2\theta}$	$k_3=\frac{1.67}{w_3}$	$\phi_m=-\frac{\beta^2}{8\times 57.3}(1+\cos 2\theta)^2\cdot\cot 2\theta$
						(B) $\int_{-\infty}^{\phi}\frac{e^{k_{IV}\zeta}d\zeta}{1+k_3^2(\phi-\zeta)^2}$		$k_3=\frac{2}{w_3}$	
W_V	Receiving slit	$f_4(\phi)$	$f_5(\phi)$	w_4	w_5	$f_5(\phi)=\int_{-\infty}^{+\infty} W_V(\zeta)f_4(\phi-\zeta)d\zeta$ (A) $\int_{\phi-w_5/2}^{\phi+w_5/2}\exp[-k_4^2\zeta^2]d\zeta$		$k_4=\frac{1.67}{w_4}$	
						(B) $\int_{\phi-w_5/2}^{\phi+w_5/2}\frac{d\zeta}{1+k_4^2\zeta^2}$		$k_4=\frac{2}{w_4}$	
W_{VI}	$K\alpha$ doublet	$f_5(\phi)$	$f_6(\phi)$	w_5	w_6				

* ϕ = angular displacement from the theoretical diffraction angle, 2θ.
$f(\phi)$ = pure diffraction contour.
w = width at half maximum intensity of the pure diffraction contour.
(A) Form of the convolution expression for a $\exp[-(k^2\phi^2)]$ contour.
(B) Form of the convolution expression for a $1/(1+k^2\phi^2)$ contour.

to 1.40. When powders of moderate or large absorption coefficients are used, the broadening ratio for the old Norelco spectrometer using CuKα radiation is seldom in excess of 1.20 and it often is less than 1.10. Even so, this instrumental factor is more important, in general, than the flat sample or vertical divergence factor in causing contour broadening.

The present analysis of sample penetration has presupposed the use of an infinitely thick sample. This condition is effectively realized under spectrometer conditions if the sample thickness, t, is great enough to satisfy the condition[12]

$$t \geq \frac{3.2}{\mu} \cdot \frac{\rho}{\rho'} \sin\theta, \tag{36}$$

in which μ and ρ are respectively the linear absorption coefficient and density of the solid material composing the powder, and ρ' is the density of the powder including interstices. The contour distortion due to sample penetration can be minimized if necessary in special cases, but at the expense of reduced intensity, by mounting a very thin layer of powder on a suitable flat substrate.

VIII. EFFECT OF RECEIVING SLIT WIDTH (W_V)

Figure 8 illustrates diagrammatically the effect of the receiving slit upon the contour being scanned ($f_4(\phi)$ in our present notation). It is assumed that the sensitivity of all points within the slit area is uniform. When the center of the slit is set at an angle of deviation ϕ_1 from the peak maximum (ideally 2θ), the slit intercepts a portion of the contour extending between the limits $\phi_1 - w_s/2$ and $\phi_1 + w_s/2$, w_s being the angular width of the slit. The recorded intensity, $f_5(\phi_1)$, corresponding to this position is evidently proportional to the area intercepted. Hence, we can write directly

$$f_5(\phi) = K \int_{\phi - w_s/2}^{\phi + w_s/2} \exp[-(k_4^2\zeta^2)]d\zeta, \tag{37}$$

assuming $f_4(\phi)$ to be an error curve in form. If we assume it to be more closely represented by the function $1/(1+k^2\phi^2)$,

$$f_5(\phi) = K \int_{\phi - w_s/2}^{\phi + w_s/2} \frac{d\zeta}{1 + k_4^2\zeta^2}. \tag{38}$$

In Eqs. (37) and (38) the weight function $W_V(\phi)$ may be interpreted as a constant, K, between the limits $\zeta = \phi - w_s/2$ and $\phi + w_s/2$. Equation (37) can be readily evaluated by transforming the exponent to $\frac{1}{2}(2^{\frac{1}{2}}k_4\zeta)^2$ and referring to tabulated values of the area of the standard error curve between definite limits. The proper value of k_4 is deduced, as explained earlier, by Eq. (6) or (7) depending upon which form of contour, $f_4(\phi)$, is assumed. The angular equivalent of the linear slit width, w_s (cm), is given by

$$w_s \text{ (degrees)} = w_s \text{ (cm)} \times 57.3/d,$$

d being the sample-to-receiving slit distance. This angular value of w_s is used in determining the integral limits $\phi + w_s/2$ and $\phi - w_s/2$.

Figure 3(V) shows the effect of a receiving slit of width $w_s = 0.045$ cm $= 0.20°$ upon an error curve diffraction contour for which $w_4 = 0.20°$ (corresponding to $k_4 = 1.67/w_4 = 8.35$). The width of the resulting contour, $f_5(\phi)$, is $w_5 = 0.25°$, which corresponds to a broadening ratio of $w_5/w_4 = 0.25/0.20 = 1.25$. Figure 4(V) shows plots of the broadening ratio w_5/w_4 for the two cases of (a) $f_4(\phi) = \exp[-(k_4^2\phi^2)]$ and (b) $f_4(\phi) = 1/(1+k_4^2\phi^2)$.

IX. EFFECT OF UNRESOLVED Kα DOUBLET (W_{VI})

For Bragg angles to small to resolve the Kα doublet, the observed contour breadth will also be increased as a result of the superposition of the α_1 and α_2 contours. The width of either contour will be w_5 as a result of the action of the five instrumental factors (see Part VIII). The broadening due to an unresolved doublet has been thoroughly analyzed by Jones,[7] who developed a curve for correcting for this effect. For our purpose Jones' plot of B/B_0 as a function of d/B_0 can be more readily used if transformed to a plot of B_0/B as a function of d/B (w_6/w_5 against d/w_5 in the present notation). This curve is shown in Fig. 9. Here w_5 and w_6 represent respectively the contour widths before and after

TABLE III. Application of the method of convolution analysis to four representative cases:* Case (a) $D>1000$ A, $w=0°$, $w_I=0.20°$. Case (b) $D>1000$ A, $w=0°$, $w_I=0.07°$, Soller slits. Case (c) $D=300$ A, $w=0.305°$ by Scherrer formula, $w_I=0.20°$. Case (d) $D=300$ A, $w=0.305°$ by Scherrer formula, $w_I=0.07°$, Soller slits.

Instrumental weight function	Quantity tabulated	Values of the quantities listed in column 2 Case (a)	Case (b)	Case (c)	Case (d)
I. Source	w	0	0	0.305°	0.305°
	w/w_I	0	0	1.53	4.36
	w_1/w	—	—	1.34	1.05
	w_1	0.200°	0.070°	0.409°	0.320°
II. Flat Sample	w_1/ϕ_m	5.89	2.06	12.0	9.4
	w_2/w_1	1.01	1.06	1.00	1.00
($\phi_m = -.034°$)	w_2	0.202°	0.074°	0.409°	0.320°
III. Vertical Divergence	w_2/ϕ_m	2.77	—	5.6	—
	w_3/w_2	1.033	—	1.016	—
($\phi_m = -.073°$)	w_3	0.209°	0.074°	0.416°	0.320°
IV. Sample Penetration	k_{IV}/k_3	4.55	1.61	8.2	5.81
	w_4/w_3	1.035	1.20	1.02	1.05
	w_4	0.216°	0.089°	0.424°	0.336°
V. Receiving Slit	w_4/w_s	2.46	1.01	4.82	3.82
	w_5/w_4	1.045	1.25	1.015	1.035
	w_5	0.226°	0.111°	0.430°	0.348°
VI. Kα Doublet	d/w_5	0.354	0.721	0.186	0.230
	w_6/w_5	1.135	1.36	1.04	1.06
	w_6	0.254	0.151°	0.447°	0.369°

* D refers to the crystallite size. Other conditions assumed are: $\mu = 40$, $2\theta = 30°$, $\lambda = 1.542$A (CuKα), $w_s = 0.088°$.

[12] L. Alexander and H. P. Klug, Anal. Chem. **20**, 886 (1948).

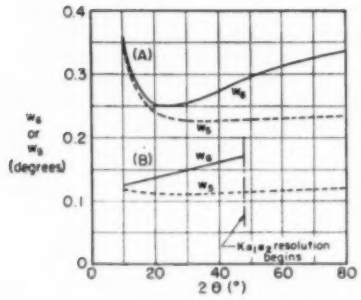


FIG. 10. Predicted dependence of contour breadth upon 2θ for spectrometers with x-ray source widths of (A) 0.20° and (B) 0.07° 2θ. w_s = width of a contour not resolving the Kα doublet; w_0 = width of a resolved line of the doublet. Conditions assumed are: $\mu=40$, $w_s=0.088°$, crystallite size greater than 1000 A.

broadening due to the unresolved doublet, and d is the angular separation of the Kα doublet characteristic of the particular radiation and diffraction angle involved (see *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, Vol. II, p. 588 ff.).

X. APPLICATION TO REPRESENTATIVE EXPERIMENTAL CASES

Table II summarizes the results of the analyses presented in the preceding sections. The five instrumental weight functions are presented in the order in which it is logical to apply them when making use of the broadening curves given in Fig. 4. Although not strictly an instrumental factor, the effect of Kα doublet broadening is included (W_{VI}) because of the necessity of taking it into account in arriving at the actually observed contour breadth at low angles. Table III illustrates with numerical data the application of the analysis to four representative experimental cases. In each case it is assumed that $\mu=40$ (moderate absorbing power for x-rays), $2\theta=30°$, $\lambda=1.542$A (CuKα radiation), and the angular width of the receiving slit, w_s, is 0.088° (a relatively narrow slit). The four specific cases carried through in the four columns of the table are:

(a) Crystallite size>1000A, $w\to 0°$, $w_I=0.20°$ (older spectrometer).

(b) Crystallite size>1000A, $w\to 0°$, $w_I=0.07°$ (newer spectrometers with narrow source and Soller slits).

(c) Crystallite size 300A, $w=0.305°$ by Scherrer formula, $w_I=0.20°$.

(d) Crystallite size 300A, $w=0.305°$ by Scherrer formula, $w_I=0.07°$.

A few important observations can be made. With a relatively wide x-ray source contour (Cases (a) and (c)) and assuming the use of a narrow receiving slit, the dominant instrumental factor causing broadening is the source width. All the other instrumental factors are inappreciable or nearly so, and the Kα doublet becomes significant only for the larger crystallite sizes. In particular, the flat sample and vertical divergence effects are very slight. These become greater at small diffraction angles since the contour widening increases with the absolute magnitude of ϕ_m for both effects (see Eqs. (17) and (25)), although flat sample broadening diminishes again when $\sin\theta$ becomes less than $\alpha d/l$ (see Part V).

The same remarks apply to the case of a narrow source except that the source is relatively less significant and the other factors more significant as causes of contour broadening. Except at low Bragg angles the flat sample and vertical divergence effects remain very minor. Penetration of the sample and Kα doublet separation both become important at the angle considered ($2\theta=30°$), particularly for contours displaying little crystallite size broadening.

In Fig. 10 the convolution analysis is applied to the problem of predicting the observed contour breadth as a function of 2θ (A) for the old Norelco spectrometer with a source width of 0.20° 2θ and (B) for newer spectrometers with an assumed source width of 0.07° 2θ and Soller slits to eliminate vertical divergence. The curves are plotted for the typical case of large crystallite size, moderately absorbing sample ($\mu=40$), and a relatively narrow receiving slit width of 0.088° 2θ. It is also assumed that the length of sample is sufficient to receive the entire beam at the lowest angle indicated ($2\theta=10°$). It will be noticed that resolution of the Kα doublet is obtained below 50° for the newer design whereas in the older design the doublet remains unresolved through the entire angular range. The failure of Curve B to make a sharp upturn below 20° such as occurs for Curve A is largely due to the absence of vertical divergence.

ACKNOWLEDGMENT

The author is appreciative of the interest and encouragement contributed by Dr. Harold P. Klug during the conduct of this investigation. He wishes to thank Mr. Douglas Pitman for valuable help in obtaining data and in performing calculations. He is also indebted to Dr. John R. Bowman for aid in evaluating the integral of Eq. (32).

Determination of Crystallite Size with the X-Ray Spectrometer*

LEROY ALEXANDER AND HAROLD P. KLUG
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received July 25, 1949)

The accuracy of the Scherrer crystallite size equation is limited in part by the uncertainty in β, the experimentally deduced pure diffraction broadening. Currently used procedures for estimating β from the observed breadth of a Debye-Scherrer line are not, in general, applicable to the x-ray spectrometer.

By making use of a scheme of convolution analysis for analyzing the effect of geometrical factors in broadening the pure diffraction contour, a correction curve is developed for determining β from the experimentally measured line breadths b and B (Jones' notation). The degree of reliability of this correction procedure is ascertained by applying Stokes' direct Fourier transform procedure for determining the form of the pure diffraction contour free of instrumental effects.

Suggestive procedures are given for crystallite size determination with the x-ray spectrometer in different size ranges, and several examples are described.

I. INTRODUCTION

THE accuracy of the familiar Scherrer equation,

$$D=(K\lambda)/(\beta\cos\theta), \tag{1}$$

is limited by the uncertainties in K, the crystallite shape factor, and β, the pure diffraction broadening. Theoretical work by Stokes and Wilson[1] has elucidated the relationship between K and the crystal shape to a degree not previously attained. However, for the usual applications to crystallite size measurement a large degree of indeterminancy in K, as much as 20 percent or even more, remains, and it is customary to let $K=1$. To gain optimum accuracy it remains to reduce the uncertainty in β as much as possible, preferably to a value much less than 20 percent.

Considerable effort has been expended by a number of investigators in evolving expressions for readily deducing β from B, the observed breadth, for the case of Debye-Scherrer lines. The most widely applied are the procedures due to Warren[2] and Jones.[3] These published correction curves have also been applied to the analysis of spectrometer maxima, although no one has specifically demonstrated their applicability to the particular conditions governing the generation of spectrometer contours. The formula of Warren,

$$\beta^2=B^2-b^2, \tag{2}$$

where b is the observed breadth of a line produced by crystallites of sufficient size to give no diffraction broadening, applies only to the case in which the β and b contours are both true "error" curves of the form $e^{-k^2x^2}$. Similarly Jones' two correction curves (a) and (b) apply only to the cases where the β contour has the respective forms $e^{-k^2x^2}$ and $1/(1+k^2x^2)$. The form of the b contour assumed by Jones was experimentally obtained by microphotometering a Debye-Scherrer line at a Bragg angle large enough to clearly resolve the Kα doublet.

Recently Shull[4] and Stokes[5] have developed Fourier transform procedures for deducing the β contour from observed b and B contours. The method of Stokes is the more generally applicable but it requires a means for the rapid summation of Fourier series. More recently one of the present authors[6] has shown how the convolution, or *Faltung*, theorem can be used to synthesize the form of a b or B contour generated by an x-ray spectrometer starting with a pure diffraction contour of known dimensions. In the light of these developments it seemed that this was a propitious time to examine the problem of crystallite size determination with the x-ray spectrometer with a view to improving the accuracy with which the width of the pure diffraction contour can be deduced free of the various instrumental broadening effects.

Since most of the commercial spectrometers in use today are of the older Norelco design which is limited to the θ range from 0° to 45°, it is felt desirable to concentrate the present analysis on crystallite size determinations which can be carried out in this angular range. Hence, we shall consider size measurements mainly in the range 0 to 500A and particularly those between about 250 and 500A, where the errors due to improper correction of the instrumental contributions may be very large. With the older type of spectrometer, having a wide x-ray source contour and limited to angles below $\theta=45°$, it is futile to attempt size measurements for crystallites much larger than 500A because the mandatory precision of the experimental measurements becomes too great for practical work.

II. COMPUTATION OF A CORRECTION CURVE FOR CRYSTALLITE SIZE DETERMINATION WITH THE X-RAY SPECTROMETER

In reference 6 the broadening effects of the five significant instrumental factors, or weight functions,

* Presented in part before the American Society for X-Ray and Electron Diffraction at Cornell University, Ithaca, New York, on June 25, 1949.

[1] A. R. Stokes and A. J. C. Wilson, Proc. Camb. Phil. Soc. **38**, 313 (1942); *ibid.* **40**, 197 (1944).

[2] B. E. Warren, J. App. Phys. **12**, 375 (1941).

[3] F. W. Jones, Proc. Roy. Soc. **A166**, 16 (1938).

[4] C. G. Shull, Phys. Rev. **70**, 679 (1946).

[5] A. R. Stokes, Proc. Phys. Soc. **61**, 382 (1948).

[6] L. Alexander, J. App. Phys. **21**, 126 (1950).

TABLE I. Data utilized in the preparation of the spectrometer correction curve. $\theta=30°$, $w_I=0.20°$, $w_s=0.088°$. By convolution analysis $b=0.210°$.

D(A)	β(°) (from Scherrer formula)	B(°) (from convolution theory)	b/B	β/B
∞	0	0.210(b)	—	—
1000	0.102	0.272	0.773	0.375
700	0.1455	0.300	0.700	0.485
500	0.204	0.341	0.616	0.598
300	0.341	0.444	0.473	0.769
200	0.510	0.596	0.352	0.856
100	1.020	1.061	0.198	0.962
50	2.040	2.040	0.103	1.000

upon a pure diffraction contour were investigated. A scheme of convolution analysis was employed to develop generalized broadening curves for each of the factors, and it was shown that by successively employing each of these curves the width of the diffraction contour generated by the spectrometer could be predicted, starting with a pure diffraction contour of any assumed width.

In reference 6 it was pointed out that with a proper choice of the experimental conditions three of the five instrumental factors causing contour broadening become negligible and only the effects of (I) the x-ray source width and (V) the receiving slit width need be considered. The requisite conditions are that the linear absorption coefficient of the powder for the x-ray beam be not too small ($\mu \geqq 40$) and that the diffraction angle 2θ be 20° or greater. The first condition is met in the case of CuKα radiation for nearly all materials except organic and metal-organic compounds, and the second condition can nearly always be satisfied by the proper choice of a diffraction line. It will accordingly be assumed in the development to follow that these simplifying conditions obtain.

The results of the convolution analysis can be easily plotted as a correction curve in the manner given by Jones[3] wherein β/B is plotted as a function of b/B. In order to simplify the terminology the notation of Jones will be used as far as possible. However, as it will prove necessary to refer to the different symbols of reference 6 occasionally, it will be helpful at the outset to compare the two notations:

Function	Jones	Reference 6
Pure diffraction breadth	β	w
Observed breadth for large crystallites (uncorrected for Kα doublet)	b_0	w_0
Same (corrected for Kα doublet)	b	w_0
Observed breadth for small crystallites (uncorrected for Kα doublet)	B_0	w_0
Same (corrected for Kα doublet)	B	w_0

It will be noticed that the contour breadths for small and large crystallites have not been distinguished in the notation of reference 6. This is because the analysis of reference 6 was of a general nature, applicable to contours of any breadth. Unless otherwise indicated the term *width*, or *breadth*, of a contour in the present paper denotes the width at half maximum intensity. Jones' analysis and correction curves apply to the integral breadth, which is defined by $\beta_x=(\int I_x dx)/(I_x)_{max}$.

The present analysis will deal mainly with the case of a relatively wide x-ray source (appropriate for the older Norelco spectrometer). For materials with crystallite size larger than 1000A the observed contour breadth b_0 will be due largely to the x-ray source and to a lesser extent to the receiving slit (unless an abnormally wide slit is employed). For materials smaller than 1000A the observed contour breadth B_0 will also be widened because of the width of the pure diffraction contour, β. If the crystallite size is very small the pure diffraction broadening becomes the dominant factor in determining the breadth of the observed contour. In addition to these effects, if the Kα doublet is not resolved, the observed contour will in all cases be widened as a result of the superposition of the overlapping contours produced by the α_1 and α_2 components of the radiation.

It will be assumed in the present paper that all observed contours have been corrected for Kα doublet broadening before the crystallite size correction curve is to be employed. The curve of Jones[3] for correcting for Kα doublet broadening is satisfactory for use with x-ray spectrometers having wide sources (of the order of 0.20° 2θ).

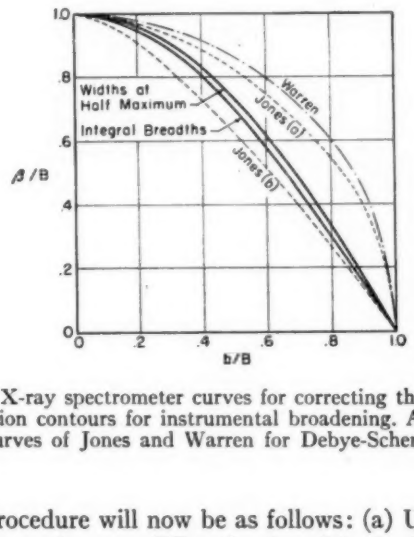


FIG. 1. X-ray spectrometer curves for correcting the breadths of diffraction contours for instrumental broadening. Also shown are the curves of Jones and Warren for Debye-Scherrer photographs.

Our procedure will now be as follows: (a) Using Eq. (1) compute the pure diffraction breadth, β, of the contour produced by a material of crystallite dimension D. (b) Using the convolution method of reference 6 compute the breadth B of the contour generated by the instrument without the added effect of the Kα doublet. (c) Similarly compute the breadth b of the contour generated for materials of large crystallite size (for which $\beta \to 0$). (d) Following the appropriate scheme of Jones, plot β/B as a function of b/B.

The success of this method of arriving at the correction curve will be contingent upon our correct choice of mathematical functions to describe the pure dif-

fraction contour and the x-ray source contour. A careful appraisal of these functions will be found in Parts III and IV of reference 6, so that it is necessary here only to repeat the final results, *viz.*, that a typical crystallite size distribution tends to generate a pure diffraction contour which is closely approximated by $1/(1+k^2\phi^2)$, whereas an x-ray source contour is more nearly described by $e^{-k^2\phi^2}$. Accordingly, the derivation of the correction curve which follows will be based on the premise that these simple mathematical functions accurately describe the pure diffraction and x-ray source contours.

The method of determining the curve of β/B as a function of b/B will be illustrated in detail by the calculation of one selected point, after which the complete data required for locating the entire curve will be tabulated. We shall select the case $\theta=30°$ and a crystallite size of $D=500$A. The spectrometer is assumed to be of the older design with an x-ray source width (w_I) of 0.200°, and a relatively narrow receiving slit width, $w_s=0.088°$, is employed. The radiation is CuKα, for which $\lambda=1.542$A.

Using the standard crystallite size equation with K set equal to unity, we find for the theoretical breadth of the pure diffraction contour

$$\beta=w=\frac{K\lambda}{D\cos\theta}=\frac{1.542\times57.3}{500\times.866}=0.204°.$$

The ratio of the width of the contour to that of the source is $w/w_I=0.204°/0.200°=1.02$. Referring to Fig. 4(I) of reference 6, we find for the broadening ratio $w_1/w=1.62$. Then the width of the resulting contour is

$$w_1=w\times(w_1/w)=0.204°\times1.62=0.330°.$$

This contour will be intermediate between $e^{-k^2\phi^2}$ and $1/(1+k^2\phi^2)$ in form because it is the convolution of two functions which are of these two pure forms and of approximately equal width.

We turn next to the action of the receiving slit in broadening this contour. The ratio of its width to that of the slit is $w_1/w_s=0.330°/0.088°=3.75$. Since we are neglecting the instrumental weight functions II, III, and IV, this ratio becomes the abscissa, w_4/w_s, of Fig. 4(V) of reference 6. The initial contour will be broadened by an amount which is approximately the mean of the values given by curves (a) and (b), hence $w_5/w_4=(1.030+1.036)/2=1.033$. This leads to a resulting contour width of

$$B=w_5=w_4\times(w_5/w_4)=0.330°\times1.033=0.341°.$$

For the reference material of large crystallite size the pure diffraction breadth β is zero. Hence, we start with the source contour and determine the broadening ratio due to the action of the receiving slit upon it. We have then

$$w_I/w_s=0.200°/0.088°=2.27,$$

TABLE II. Calculation of the crystallite sizes of four materials from spectrometer measurements after determining β (A) by Stokes' Fourier method and (B) from the correction curve of Fig. 1 (widths at half maximum).

	CaF_2	MgO	Cristobalite	Basic calcium phosphate
Observed breadths				
B_0	0.500°	0.700°	0.470°	0.468°
b_0 (quartz powder)	0.330°	0.330°	0.320°	0.330°
Breadths after correcting for $K\alpha_1\alpha_2$				
B	0.460°	0.674°	0.462°	0.456°
b	0.276°	0.284°	0.307°	0.314°
Correction ratios				
b/B	0.600	0.421	0.665	0.689
β/B (Fig. 1)	0.620	0.810	0.533	0.500
Pure diffraction breadths				
β (Stokes' method)	0.335°	0.530°	0.260°	0.215°
β (Fig. 1)	0.285°	0.546°	0.246°	0.228°
Calculated crystallite sizes				
D(A.) (Stokes' method)	290	179	346	422
D(A.) (Fig. 1)	342	174	366	398

and from Fig. 4(V), curve (a), of reference (6), $w_5/w_4=1.05$, and consequently

$$b=w_5=w_4\times(w_5/w_4)=0.200°\times1.05=0.210°.$$

From these calculated values of β, b, and B we obtain

$$b/B=0.616 \quad \text{and} \quad \beta/B=0.598.$$

By proceeding in a similar manner for a number of crystallite sizes ranging from 50 to 1000A the remaining data of Table I of the present paper were obtained. When plotted with β/B as a function of b/B, the solid curve marked "Widths at Half Maximum" of Fig. 1 is obtained. Although deduced specifically for the case of $\theta=30°$ and a receiving slit width of 0.088°, this curve is found to be perfectly general for all values of θ in the range 0–45° and for receiving slit widths ranging from 0 to about 0.20°. When the slit width exceeds the angular width of the source by an appreciable amount, the correction curve shifts to the right. The most satisfactory policy in this regard is to always employ the narrowest receiving slit that is compatible with the diffracted energy available.

By a parallel set of calculations it can be shown, or by direct comparison of the geometrical conditions it can be inferred, that this same correction curve applies, within limits, to spectrometers with narrower sources. However, when the source width becomes much smaller than 0.10°, the several minor broadening factors become relatively significant, so that the premises upon which the present curve is based are no longer correct.

III. EXPERIMENTAL SUBSTANTIATION OF THE CORRECTION CURVE

In Part III of reference 6 the pure diffraction contours of six miscellaneous materials were investigated, five of them by the Fourier transform method of Stokes.[5] Four of these materials made excellent subjects

TABLE III. Comparison of the crystallite sizes of four materials as deduced by different correction curves. The figures are calculated crystallite sizes in A.

Method of correction	CaF_2	MgO	Cristobalite	Basic calcium phosphate
Stokes (Fourier)	290	179	346	422
Curve of Fig. 1, widths at half maximum	342	174	366	398
Jones' curve (a)	285	162	284	301
Jones' curve (b)	414	201	448	491
Warren's curve $[\beta/B=(1-b^2/B^2)^{\frac{1}{2}}]$	265	156	260	274

for a test of the spectrometer correction curve of Fig. 1. The widths of the pure contours deduced by the Fourier method can be considered to be the best obtainable and hence to also lead to the most reliable values of the crystallite sizes. In Table II the crystallite sizes are calculated using the curve of Fig. 1 marked "Widths at Half Maximum," and the resulting values are compared with the presumably best values from the Stokes procedure.

In every case an appropriate quartz maximum was measured in order to obtain the required value of b_0 for a material of large crystallite size. The contours were measured with a Norelco Geiger-Counter Spectrometer, Type No. 12021 (with wide x-ray source), the spectrometer arm being set manually at appropriate intervals and the counting rates being recorded directly. In every instance except that of basic calcium phosphate it was possible to reduce the probable statistical counting error in the peak height above background to 1 percent or less. It is seen that in three out of four cases the agreement is highly satisfactory, and, furthermore, that in the single case of less satisfactory agreement (CaF_2) a greater deviation from the presumably best value (by Stokes' method) is to be expected because the pure diffraction contour is intermediate in nature between the forms $1/(1+k^2\phi^2)$ and $e^{-k^2\phi^2}$ rather than resembling more closely the first function (refer to Table I of reference 6).

Table III compares the crystallite sizes of these same materials as calculated using various correction curves with the values arrived at by Stokes' Fourier method. For the sake of ready comparison these correction curves are also plotted in Fig. 1. An examination of Table III shows that on the average the present curve gives much better agreement with the Fourier results than any of the other curves. Only in the case of the somewhat anomalous contour of CaF_2 would somewhat better results have been obtained by using other curves, e.g., by employing either Warren's curve or Jones' curve (a). It is interesting to note from Fig. 1 that the new spectrometer curve agrees more nearly with Jones' curve (b) at large values of b/B (large crystallite sizes), while it approaches both Jones' curve (a) and Warren's curve more closely at small values of b/B (small crystallite sizes). However, as is well-known, the choice of correction curve is unimportant for small crystallite sizes.

IV. DISCUSSION OF THE CORRECTION CURVE

In Part II it was emphasized that the derivation of the present correction curve is based on the premise that the functions $1/(1+k^2\phi^2)$ and $e^{-k^2\phi^2}$ accurately describe the pure diffraction contour and x-ray source contour respectively. These formulas should be regarded realistically as being only limiting functions which are approached only more or less imperfectly in any experimental situation. Consequently the most that should be claimed for the present correction curve is that the probability of its yielding an accurate value of β is good and also much better than that of curves based on other pairs of simple mathematical functions.

In particular, it is suggested that each investigator test the focal spot of the particular x-ray tube employed for agreement with the function $e^{-k^2\phi^2}$. A simple but sufficiently accurate method for accomplishing this is described in Part IV of reference 6. If reasonably good agreement is found he may then use the present correction curve with confidence.

For a spectrometer with a wide source the correction curve of Fig. 1 will have rather general reliability provided that the following experimental requirements are satisfied: (a) The linear absorption coefficient of the powder (including solid and interstices) should be approximately 40 or larger. (b) The diffraction angle, 2θ, should be greater than 20°. (c) The receiving slit width should not exceed the width at half maximum intensity of the x-ray source contour (both expressed in terms of 2θ units). (d) By proper adjustment of the x-ray slit the horizontal and vertical divergence of the x-ray beam should preferably be limited to moderate values, say, 1° and 2° respectively. Under these circumstances, if we disregard the usual remaining uncertainty in K of Eq. (1), satisfactory measurements can be made of crystallite sizes as large as 500A provided that B_0 and b_0 can be measured with an accuracy of ±0.01° or better for the larger sizes. For the new spectrometers with narrower x-ray sources the present curve is fairly reliable for crystallite size measurements below 500A. However, although instruments of this design are theoretically capable of application in the size range 500-1000A, the novel geometrical factors involved are such as to require a special analysis in order that a correction curve may be established for work in this size range.

If it is desired to use integral contour breadths, defined by $\beta_z=(\int I_z dz)/(I_z)_{\max}$, rather than widths at half maximum intensity, the curve marked "Integral Breadths" in Fig. 1 should be employed. This curve was derived by substituting the functions $1/(1+k_1^2x^2)$ and $\exp(-k_2^2x^2)$ respectively for $F(kx)$ and $f(x)$ in Jones' Eqs. (4) and (5).[3] The integral breadth is more convenient for mathematical analysis but less readily employed in the experimental measurement of crystallite size.

When measurements of the highest reliability are required, it is recommended that the Fourier transform method of Stokes be employed for determining β. Once the equipment for the rapid summation of Fourier series has been constructed and the technique for performing the operations has been mastered, the time consumption per analysis is not excessive (possibly one or two hours). The Fourier method has completely

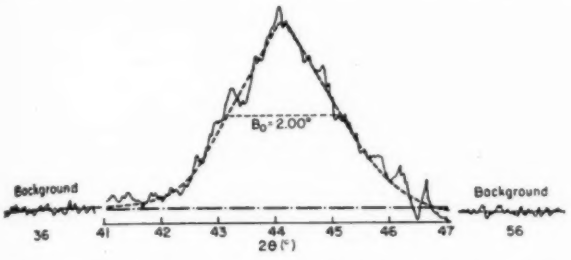


FIG. 2. Determination of the crystallite size of a nickel catalyst from an automatic strip chart record of the (111) maximum. CuKα-radiation; scanning speed 1° per minute; background measured at $2\theta=36°$ and 56°. Neglecting b_0, $D_{111}=48$ A. Using $b_0=0.33°$ for the (220) reflection of NaCl powder, $D_{111}=49$ A.

general applicability regardless of the particular forms of the pure diffraction contour and the various instrumental broadening functions. Furthermore Kα doublet broadening is automatically allowed for.

V. PRACTICAL SUGGESTIONS FOR CRYSTALLITE SIZE DETERMINATION WITH THE X-RAY SPECTROMETER

The following discussion applies to the older spectrometer design incorporating a wide x-ray source and a working range of $2\theta=0$ to 90°. Three crystallite size ranges will be given special consideration, each amenable to a somewhat different experimental approach.

For all except the smallest crystallite sizes it is necessary to measure the contour widths of both the material being studied (B_0) and a reference material which gives no crystallite size broadening (b_0). In order that the geometrical broadening effects may be as nearly identical as possible, the two peaks should be situated at approximately the same diffraction angle. When the linear absorption coefficients for x-rays of the "unknown" and reference powders are both high (say, $\mu=40$ or more when the interstices of the powder are included in the calculation) or when the absorption coefficients are known to be very nearly equal, good results can be obtained when the two materials are mounted and analyzed separately. However, when the coefficients are small, and in particular when they are known to differ considerably, the two materials should be intimately mixed and the measurements made on a single sample. The fact that the calculated absorption coefficients of the two solid materials are nearly equal is in itself no guarantee that the powders will absorb equally because they may not pack to the same density. When this is evidently true, or when there is any doubt in this regard, the materials should always be mixed. Fine quartz powder manifests a fairly typical packing behavior and is found to assume a state in which solid and interstices constitute about 40 and 60 percent respectively of the total sample volume. Hence, the linear absorption coefficient of the powder is only 0.40 as large as that of the solid material [for CuKα radiation μ (quartz powder) $=0.40\times93=37$].

The above suggestions presuppose the use of a sample of sufficient thickness to give maximum absorption (see remarks in reference 6, Part VII). If sufficient intensity can be obtained with the use of a very thin sample, such as a film of powder mounted on a suitable plane substrate, the differences in absorption become much less important and it is usually unnecessary to mix the powders.

Case 1. Size Range < 150A

In this range fairly good results can be had by preparing automatic strip chart records at a sufficiently slow scanning speed to yield a contour of easily measurable width. The background is also measured at selected points on either side of the maximum, after which the best straight horizontal line is drawn through the statistical fluctuations of each background record and a smooth curve is drawn to delineate the peak contour. This method is not successful unless the peaks in question are relatively strong. In favorable cases of this sort β can be determined with a probable error of 5 to 10 percent. Figure 2 illustrates the application of this method to a nickel catalyst of fine crystallite size.

If the peak contours are counted and plotted manually and if sufficient counts are taken to permit the

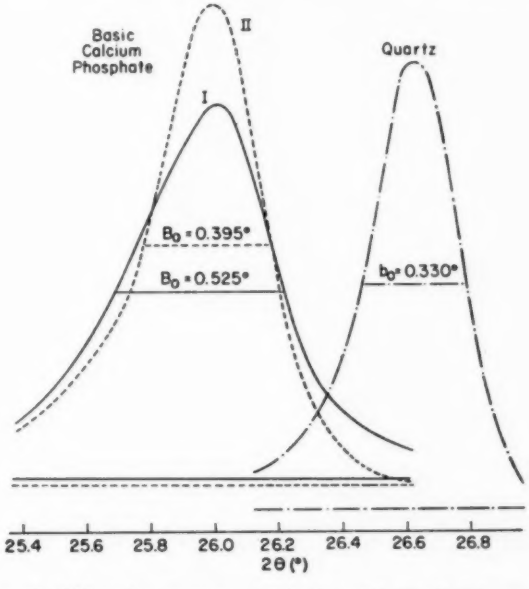


FIG. 3. Diffraction contours of basic calcium phosphate precipitate before (I) and after (II) digestion, and reference contour of quartz powder. The contours were plotted from manually recorded counting rate data. Crystallite sizes (after correcting for Kα doublet broadening): (I) 290 A, (II) approximately 800 A.

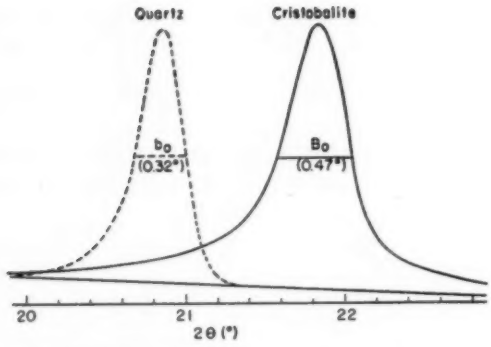


FIG. 4. Diffraction contours used in determining the crystallite size of cristobalite prepared by calcining diatomaceous earth. The contours were plotted from manually recorded counting rate data. Mean crystallite size (after correcting b_0 and B_0 for Kα doublet broadening), 366 A.

determination of the peak height above background with a probable error from counting statistics of 1 percent or less (5000 counts per point for cases with negligibly small backgrounds), the probable error in β can be reduced to 3 percent or less. In this size range it is unnecessary to correct the observed line breadths for Kα doublet broadening. Warren[2] has pointed out that the measured breadth of the "unknown" line, B_0, can be substituted for β directly in the crystallite size formula when $b_0 < B_0/5$, because in such instances the broadening due to the instrumental effects becomes insignificant in comparison with the pure diffraction broadening. Thus, for the nickel catalyst of Fig. 2 the calculated crystallite size was 49A when b_0 was taken into account and 48A when not.

Case 2. Size Range 150–300A

Automatic recording is not satisfactory in this range except for very rough measurements. If the heights of the peaks above background are determined with an accuracy approaching 1 percent (as limited by counting statistics), it is possible to determine β with a probable error of 5 percent. This and succeeding estimates of precision and error take into account a reasonable degree of uncertainty in the position of the spectrometer correction curve (Fig. 1). In this size range it is unnecessary to correct for Kα doublet broadening for $2\theta < 60°$. For higher angles the error involved is sufficient to increase appreciably the 5 percent probable error quoted above.

Case 3. Size Range 300–500A

In this range it is necessary to adopt the best experimental techniques in order to secure good results. A probable error in β of considerably less than 10 percent can be attained by keeping the probable error in counting statistics at points near the tops of the peaks smaller than 1 percent and by correcting for Kα doublet broadening *in all cases*. In addition, for the most accurate work it is recommended that the recorded counting rates be corrected for the nonlinearity of response of the Geiger tube.[7]

Of course, care must also be exercised in defining accurately the sides of the contours at the levels at which the breadths B_0 and b_0 are to be measured. This means measuring the counting rates at two or more points in the immediate vicinities. An analysis of the counting errors involved has shown that it is satisfactory to use a constant time interval for counting all the points defining a given contour, provided only that this time interval furnishes a sufficient number of counts to determine the height of the peak above background with high precision.

The background should preferably be counted with a longer time interval because of the normally much smaller intensities involved, and it is suggested that this interval be made twice as long as that employed in counting the peak contour. When the background is abnormally high and constitutes an appreciable portion of the over-all counting rate measured at the top of the peak, it is necessary to totalize much more than 5000 counts at the top of the peak in order to reduce the probable error in the *net* peak height to less than 1 percent.[8] In any event care should also be taken that the background be measured at points sufficiently far removed from the peak to avoid contributions from the base of the peak itself. For sharp peaks an angular displacement from the peak center of 2.0° is usually satisfactory, but for the very wide maxima generated by a distribution of crystallite sizes in the range below 100A it is necessary to increase this to 5° or even more.

Figures 3 and 4 illustrate crystallite size measurements in the ranges of Cases 2 and 3. In each figure the required reference contour of a material of large crystallite size (fine quartz powder) is also shown. Actually each contour pictured was obtained using a separate sample, but the several contours have been drawn together to facilitate comparison. The growth of the basic calcium phosphate crystallites during digestion is strikingly reflected in the alteration of the diffraction contour shown in Fig. 3. The accuracy of the measurement following digestion is obviously not good (probably not better than 800±150A) because the size involved is outside the satisfactory working range of the spectrometer. Figure 4 portrays the b_0 and B_0 contours which were analyzed in determining the crystallite size of cristobalite, a form of crystalline silica, prepared by calcining diatomaceous earth.

ACKNOWLEDGMENT

The authors are indebted to Elizabeth Kummer and Douglas Pitman for assistance in obtaining the experimental data and performing calculations.

[7] Alexander, Kummer, and Klug, J. App. Phys. **20**, 735 (1949).
[8] Tracerlog No. 2, p. 5, March 1947.

Radiofrequency Mass Spectrometer

WILLARD H. BENNETT
National Bureau of Standards, Washington, D. C.
(Received August 29, 1949)

A mass spectrometer tube using velocity selection is described. The tube has parallel plane grids made with knitted wire nets with a large percentage of open area. The grids are arranged in groups of three, and radiofrequency alternating potential is applied to the middle grid of each group. Stopping potentials are used to turn back all ions except those with the selected mass. The spectrometer has sufficient resolution for ordinary gas analysis requirements and is simpler, more compact, and more rugged than magnetic beam deflection devices.

SEVERAL proposals have appeared in the past for making a mass spectrometer using velocity selection instead of a magnetic deflection of a focused beam of ions.

Smythe and Mattauch[1] used a succession of pairs of condenser plates along a tube, with which radiofrequency alternating fields could be applied to the ions in a collimated beam at right angles to the principal direction of movement of the ions in the beam. The ions in order to arrive at the receiver had to have the velocity necessary to place them opposite each successive pair of condenser plates at successively opposite phase. Because harmonic masses would also be recorded in such an arrangement, Smythe and Mattauch[1] also bent the beam through an arc with an electric field between a curved pair of electrodes in order to eliminate harmonics. From the available publications, this method does not appear to have been successful.

Stephens[2] proposed, and Cameron and Eggers[3] built and tried a pulsed ion source in combination with an oscilloscopic display of the distribution in time of the arrival of ions at a receiver. The resolution obtained, using this attractively simple idea, is disappointingly limited. This limitation is probably inherent in any method which has to use such high speed recording of such small pulses of current.

Weisz[4] has suggested, but apparently not tried, an adaptation of a linear accelerator in which alternately phased radiofrequency alternating potential is applied to a succession of equally spaced coaxial cylindrical electrodes. He proposes to operate the instrument in a manner such that the ions to be selected arrive at the gaps between electrodes at zero phase and that they be selected on the basis that they have neither gained nor lost energy because of the alternating fields between the successive cylindrical electrodes. He fails to make clear how to accomplish such a selection.

Bauer[5] has suggested a method which requires a well-focused beam to travel between parallel plates between which an alternating radiofrequency potential is applied. This method also appears not to have been tried.

Proposals have also appeared[6] for time-of-flight methods which use uniform magnetic fields in obtaining well-focused mass components, and as such have most of the disadvantages of the more familiar magnetic beam deflection methods and can use only as intense an ion beam as can be made to pass through a narrow slit.

The present development grew out of some observations made in 1946 on a pentode with coaxial cylindrical electrodes. A radiofrequency a.c. potential was applied to the intermediate one of the three grids while negative ions were produced at the hot cathode in the central portion of the tube. Although this tube was built for the purpose of separating kinds of negative ions differing in mass by a factor of two or more, it was found that a much better separation was effected than had been anticipated. For this reason, experiments were initiated to explore the possibilities of this kind of velocity selector as a mass spectrometer.

SINGLE-STAGE TUBE

The first tube built using parallel plane grids was a single-stage tube, as shown in Fig. 1, and illustrated schematically in Fig. 2. Electrons from the filament, *a*, are accelerated through a d.c. potential to the first grid, *b*. Positive ions produced by collisions between these electrons and gas molecules in the tube between grids *b* and *c* are accelerated from *b* toward *c* with a potential difference, *V*, between *b* and *c*. The grids *c*, *d*, and *e* are equally spaced at a distance *s*. A radiofrequency field of angular frequency ω is applied to the middle grid *d* giving an electric field between *c* and *d*

$$E_{cd}=E\sin(\omega t+\theta)$$

and between *d* and *e*

$$E_{de}=-E\sin(\omega t+\theta),$$

where *t* is zero at the instant an ion crosses the plane of grid *c* and $E\sin\theta$ is the value of the field E_{cd} at that instant. We assume a uniform flow of ions up to grid *c* so that there will be equal numbers of ions with each

[1] W. R. Smythe and J. Mattauch, Phys. Rev. **40**, 429 (1932).
[2] W. E. Stephens, Phys. Rev. **69**, 691(A) (1946).
[3] A. E. Cameron and D. F. Eggers, Jr., Rev. Sci. Inst. **19**, 605 (1948).
[4] P. B. Weisz, Phys. Rev. **70**, 91(L) (1946).
[5] S. H. Bauer, J. Phys. Chem. **39**, 959 (1935).
[6] S. A. Goudsmit, Phys. Rev. **74**, 622 (1948). J. A. Hipple and H. A. Thomas, Phys. Rev. **75**, 1616 (1949).

possible value of θ. Some of these will receive energy from the fields as they pass the grids while others will lose energy to the fields. The collecting plate f has a d.c. stopping potential which turns back all ions except those which have acquired nearly the maximum possible energy from the fields and which originated near the grid b.

An exact solution for the motion of a charged particle in the sinusoidal field is extremely complex and has not been obtained. For purposes of this discussion, a small disturbance approximation will suffice. Experiment has shown that this approximation holds quite well for finite amplitudes. It is assumed that the change in velocity of an ion as it moves through the grids is small compared to its initial velocity so that the effects

FIG. 1.

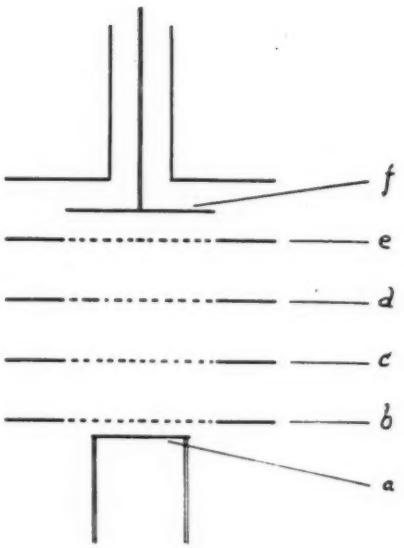


FIG. 2.

of the applied fields on the transit time can be neglected. Space-charge effects are also neglected with negligible error for this system.

The force acting on mass m between grids c and d is

$$F_{cd}=m\ddot{x}=Ee\sin(\omega t+\theta)$$

and between d and e is

$$F_{de}=m\ddot{x}=-Ee\sin(\omega t+\theta),$$

where time t is measured from the time the ion passes the grid c and θ is the phase angle of the alternating potential on d at the instant the ion passes c. Taking for a first approximation, the case when the amplitude of the alternating potential is small compared with the d.c. potential, V, used in accelerating the ion from the source to the grid, c, the variations in the velocity of the ion due to the a.c. fields will be small compared with the average velocity of the ion between grids c and e. The energy acquired from the alternating field is approximately:

$$\Delta W=\Delta(mv^2/2)=v\cdot\Delta(mv)=v\int Fdt.$$

The transit time for the ion to travel from one grid to the next is s/v where s is the distance between grids. Then

$$\Delta W=v\left\{\int_0^{s/v}Ee\sin(\omega t+\theta)dt+\int_{s/v}^{2s/v}-Ee\sin(\omega t+\theta)dt\right\}$$

$$\Delta W=\frac{Eev}{\omega}\left[\cos\theta-2\cos\left(\frac{s\omega}{v}+\theta\right)+\cos\left(\frac{2s\omega}{v}+\theta\right)\right].$$

This is a maximum with respect to variations in θ when

$$(s\omega)/v+\theta=\pi=180°,$$

showing that the ions which pass grid d just as the field reverses obtain maximum incremental energy. The expression for ΔW is a maximum with respect to variations in ω when

$$\theta = 46° \text{ (more precisely } 46°26')$$

so the transit angle between grids is

$$(s\omega)/v = 134°.$$

The velocity of the ion receiving maximum incremental energy is obtained from

$$eV = \tfrac{1}{2} M m_0 v^2,$$

where V is the potential difference between the ion source, b, and the grid, c; M is the atomic mass number of the ion; and m_0 is the mass of an ion of unit mass number. Eliminating v between the last two equations gives

$$M = \frac{0.266.10^{12} V}{s^2 f^2},$$

where V is in volts, s is in centimeters, and f is in cycles per second.

Shown in Fig. 3 is the energy, ΔW, acquired by an ion passing the grid d, at 180° phase as a function of the number of cycles, N, executed by the alternating field while the ion travels from grid c to grid e. The principal maximum of this curve is at $N = 0.74$ cycle because the transit angle from the first to the third grid for maximum ΔW is

$$(2s\omega)/v = 267° = 0.74 \text{ cycle}.$$

If the stopping potential applied in the collecting system corresponds to the energy Z, at the horizontal dashed line, the ions can be collected only at frequencies corresponding to positions in the graph between A and B. Ions passing the first grid at phase angles different from the optimum 46° angle acquire a lower energy and can reach the collecting electrode within a narrower range than the range between A and B. The ions collected are those at all phases of entry which can collect enough ΔW to exceed Z so the shape of the line observed for ions of a particular mass is sharper than that portion of the curve above Z.

An example of the line form observed is shown in Fig. 4. This is a plot of the experimentally observed ion currents for mercury positive ions. A smaller stopping potential is used for the upper curve than for the lower curve.

It was found by experiment that the above approximate treatment was adequate if the d.c. potential difference between the source and the first grid was more than 10 times the amplitude of the r-f alternating potential, but that irregular curves for each mass component resulted at larger relative a.c. potentials. However, symmetric mass line forms resulted for a.c. potentials up to about half the d.c. potential, if the d.c. potential on grid d, to which the a.c. potential is added, is reduced by just the amount necessary to reduce the speed of an ion receiving maximum incremental energy, back to the same speed as that with which the ion passed grid c, and if the d.c. potential on grid e were also similarly reduced below that of grid d.

A good example of this can be given in the case where the tube is used for resolving the negative atomic ions produced at the surface of a hot cathode. In this case the potentials applied were:

Ion source, a,	0.0 volt, d.c.
Grid, b,	+10.0 volts d.c.
Grid, c,	+ 4.0 volts d.c. plus 6 volts (r.m.s.) a.c.
Grid, d,	− 2.0 volts d.c.
Grid, e,	−14.1 volts d.c.
Collector, f	+20.0 volts d.c.

It will be noted that in this particular test, the stopping potential was applied to the fourth grid, e, and not to the collector. A simple integration shows that if the grids were ideal grids (very small holes compared with the inter-grid distance), the maximum incremental energy an ion could acquire from the r-f fields would be about the same as would be acquired from a d.c. potential twice the r.m.s. value of the r-f potential on the intermediate grid. Thus the use of an r-f potential of six volts requires stepping back the d.c. potential on each successive grid by about six volts in order to hold approximately constant the velocity of those ions receiving nearly maximum incremental energy. The value of the stopping potential, −14.1 volts, exceeds the value of a little less than −12 expected because the more positive potential of the adjacent electrodes reduces the magnitude of the potential of the middles of the holes in the stopping grid to less than that of the potential applied to that grid.

By appropriately selecting the d.c. potentials, this spectrometer tube serves equally well for separating positive or negative ions. Although no magnetic field is needed with this tube in separating positive ions, it is helpful to use a small magnetic field perpendicular to

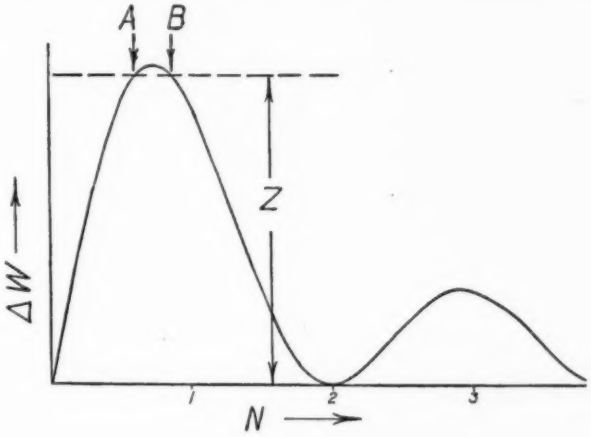


FIG. 3.

the axis of the tube when separating negative ions in order to confine the electrons from the source to the vicinity of the source. This prevents the formation of positive ions throughout the tube by the relatively large electron currents which would otherwise reach the vicinity of the stopping grid.

TWO-STAGE TUBE

The method was extended to a two-stage tube shown schematically in Fig. 5. In this case, the grids *b*, *c*, and *d* serve as the first stage while grids *e*, *f*, and *g* function as the second stage. The same radiofrequency alternating potential is applied to grids *c* and *f* in addition to the two unequal d.c. potentials whose difference provides the compensation described previously in connection with the single-stage tube. As before, the potential *V* accelerates the ions from the source *a* to the first grid *b*, and the d.c. potentials of grids *c*, *d*, *f*, and *g* are similarly adjusted. The grids *d* and *e* are at the same potential so that the region between them is an equipotential drift space. The stopping grid *h* and the collecting plate *p* function as before.

The distance between grids *d* and *e* is such that an ion receiving maximum incremental energy from the first stage will reach the initial grid of the second stage, *e*, at the right phase of the r-f alternating potential to acquire a maximum incremental energy from the second stage also. In order to accomplish this, the r-f potential must execute exactly an integral number of cycles while the ion travels from stage to stage, or more exactly, from *c* to *f*, and so the distance from *d* to *e* must be

$$(2.70n-2)s,$$

where *n* is an integer, and *s* is the distance between the grids in each stage as before. The first two-stage tube was built with $n=6$. This will be referred to as a six-cycle two-stage tube.

The incremental energy which can be acquired by an ion from both stages of a two-stage tube is illustrated in Fig. 6. The envelope of the curve is the same kind of a curve as plotted in Fig. 3 but with twice the ordinate, of course.

In Fig. 7 is shown the current of negative iodine ions from a hot tungsten filament in iodine vapor at 10^{-4}

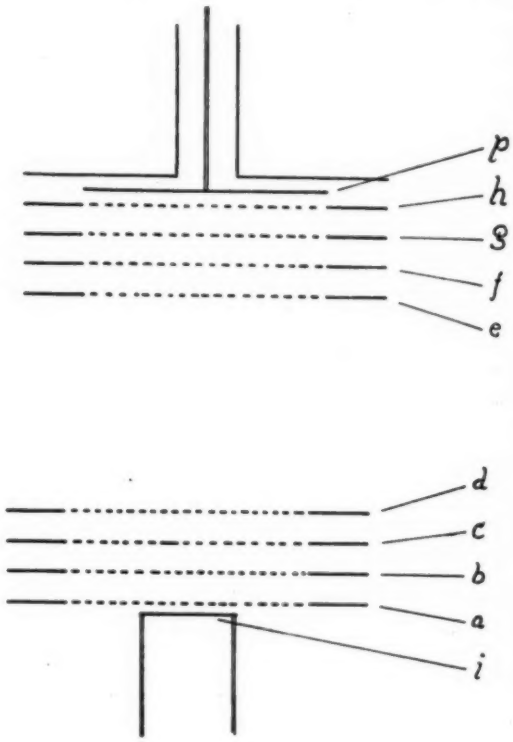


FIG. 5.

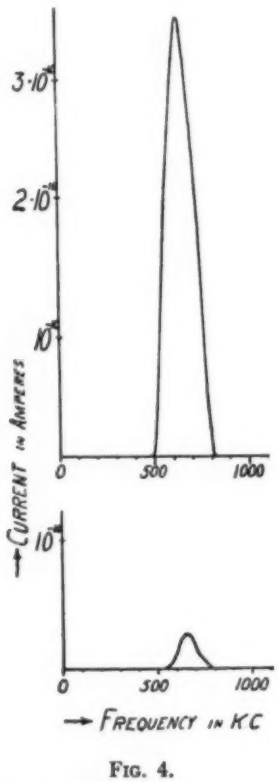


FIG. 4.

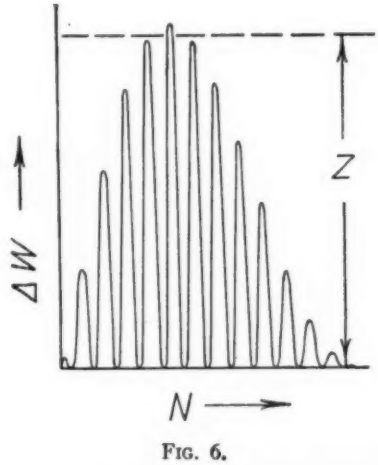


FIG. 6.

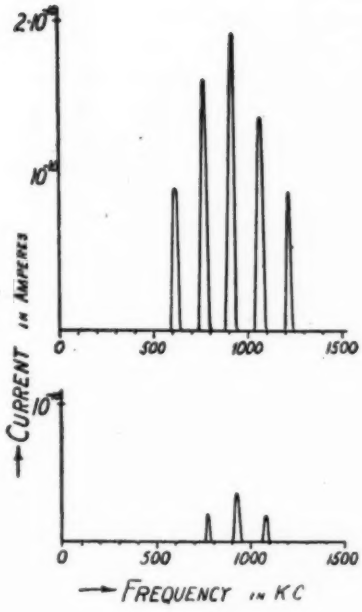


FIG. 7.

mm pressure. The lower curve shows the currents observed using a larger stopping potential, Z, than that used in obtaining the upper curve. It is seen by comparing these curves with those in Fig. 4 that the selection is much sharper than that obtained with a single-stage tube. High peaks occur for frequencies corresponding to other numbers of cycles between stages than the number 6 for the principal peak, and much of the sensitivity of the tube is lost in increasing the stopping potential enough to cut out the adjacent peaks.

THREE-STAGE TUBE

In order to retain much more sensitivity and, at the same time, produce only a single peak for each mass, a third stage is introduced. Such a tube is shown in Fig. 8. The three stages must be spaced so that the r-f alternating potential on the middle grid of each stage executes exactly integral numbers of cycles while the ion acquiring maximum incremental energy travels from stage to stage, i.e., from e to h, and from h to k. These integral numbers between the two successive pairs of stages are so selected as to avoid any appreciable overlap of the harmonic peaks of the two pairs of stages considered as two-stage combinations. This is illustrated in Fig. 9. When the integral numbers are so selected, much more of the principal peak can be left above the blocking potential, Z, than with either the single-stage or two-stage tubes and a corresponding increase in sensitivity is effected.

A 9–7-cycle three-stage tube was built, shown in Fig. 8, and illustrated schematically in Fig. 10. This tube was provided with two different kinds of ion source. One made use of a pencil of electrons, a', which passed at right angles to the axis of the tube. Ions produced along this electron beam were made to drift toward the grid c with a small electric field between b and c. The other source of ions which could alternatively be used made use of electrons from filament a accelerated toward grid b. Ions produced between b and c were similarly made to drift toward c with a small field between b and c.

A d.c. potential, V, accelerates the positive ions emerging through c toward and through d. The grids d, e, and f act as the first stage; g, h, and i the second stage; and j, k, and l the third stage. The only grids receiving the radiofrequency a.c. potential are e, h, and k. The stopping potential difference is applied between the collecting plate p and the grid c. The grid m was

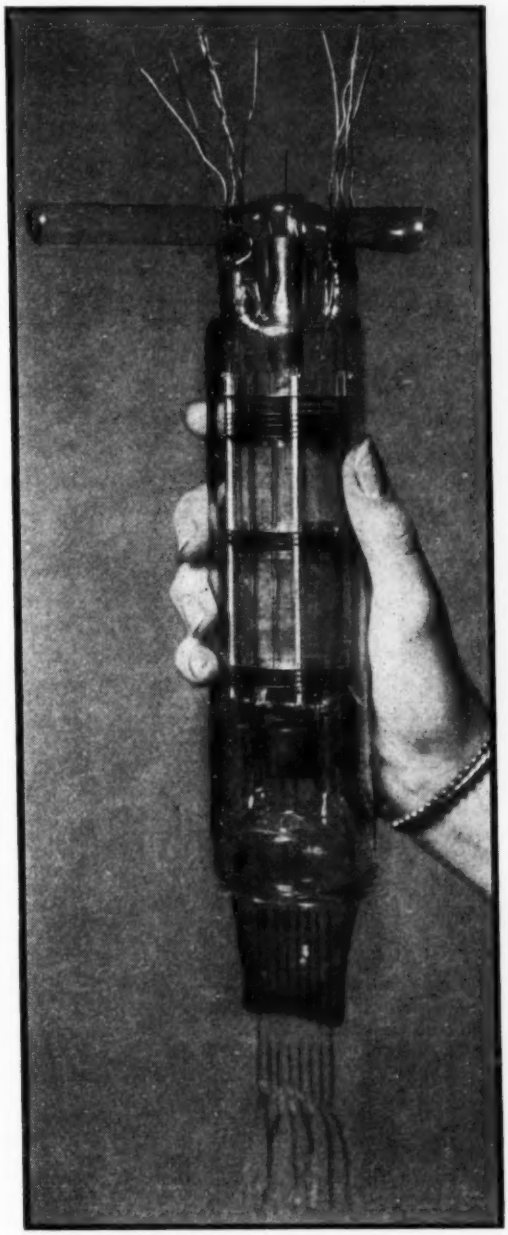


FIG. 8.

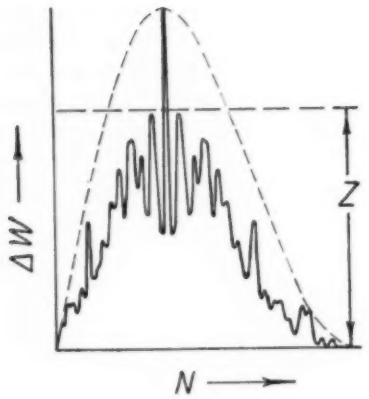


FIG. 9.

held at the same potential as the collector, and grid *n* was held at a more negative potential than any other electrode in the tube in order to turn back all electrons produced in the tube between grids *d* and *n*.

The grids used in this work were made with 0.0005-in. tungsten wire knitted with about 15 loops per centimeter. One of these grids is pictured in Fig. 11. The wire is mounted on the frame in a stretched condition so that the grid does not sag during or after baking. This is quite important in order that the grids remain accurately positioned with respect to each other. The open area of these grids is greater than 95 percent and even though 13 are used in a tube, more than 50 percent of the beam can pass through to the collector. The most open of the woven wire screens available with openings this small have open areas less than 65 percent, and 13 of these in series would pass less than one-half percent and would be quite impractical. Mr. William B. Haliday of this Bureau has assisted considerably in the development of the techniques for making these grids.

In this first of the three-stage tubes built, the grids in the stages were spaced at 2 mm which is of the same order of magnitude as the size of the holes in the grids. For this reason, the grids were far from ideal grids.

Figure 12 shows one of the observations of the spectrum of positive ions of mercury. In this observation, harmonic peaks occur at frequencies above and below the principal peak by about one-eighth the frequency of the principal peak. These harmonic peaks are evidently due to a partial overlap of the first harmonic peaks of the two two-stage combinations whose harmonics occur at frequencies, respectively, at one-ninth and one-seventh of the principal frequency, as illustrated in connection with Fig. 7.

Increasing the stopping potential and correspondingly reducing the intensity of the observed line eliminates these harmonic peaks as shown in Fig. 13. In this figure, the solid vertical lines under the curve are drawn proportional to the relative abundance of the isotopes. The width of this observed line at half the maximum height is $3\frac{1}{2}$ percent while the width, due to the spread in isotope distribution, should be at least $1\frac{1}{2}$ percent. However, the line flares out at the base much more than the lines did in the two-stage cases. This is believed to be due to the fact that this tube used grid

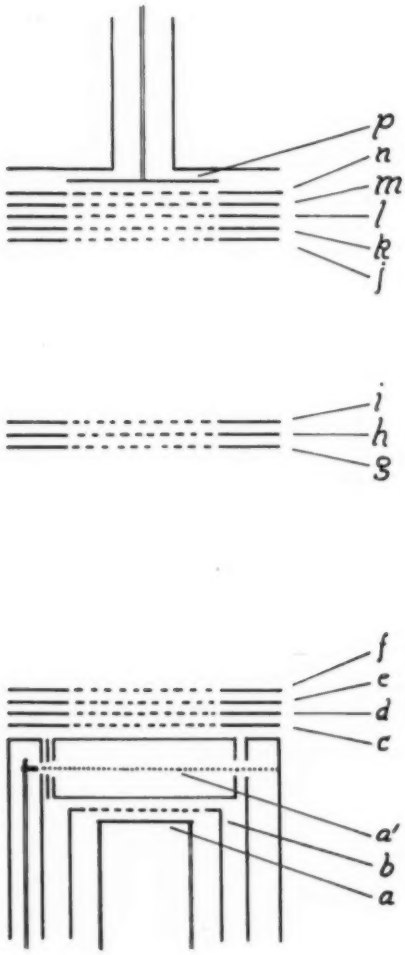


FIG. 10.

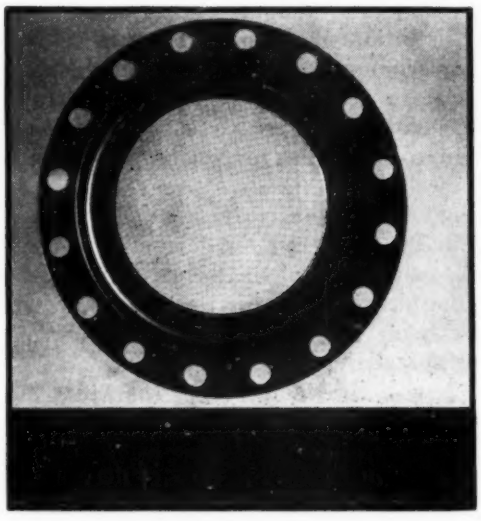


FIG. 11.

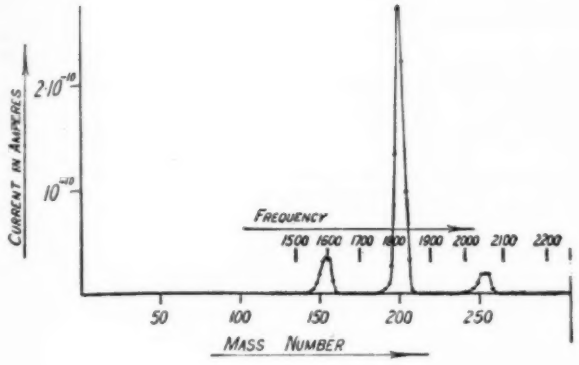


FIG. 12.

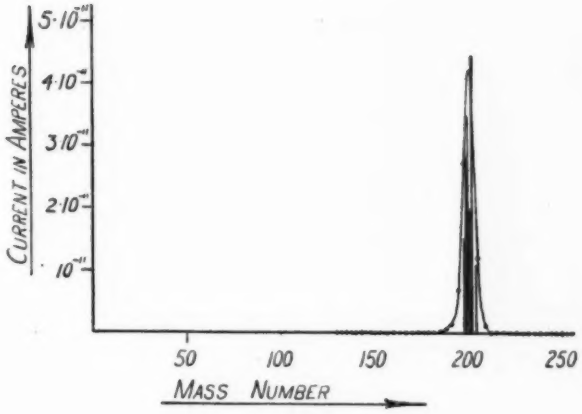


FIG. 13.

distances which were too small compared with the size of the holes in the grids.

From the shape of the line shown in Fig. 13, it is seen that this spectrometer should have sufficient resolution for those gas analyses where the masses to be completely separated differ by more than six percent, as is the case in many ordinary gas analyses of the lighter substances.

It should be noted that after constructing this spectrometer tube with proper care given to the accuracy of the grid spacings, and the accuracy of measurement of the potential, V, the formula

$$M = \frac{0.266.10^{12}V}{s^2f^2}$$

accurately gives the frequencies at which the various mass components appear. The instrument, when built with due attention to grid spacing, does not have to be calibrated with known gases and, in this sense, this is an absolute instrument.

This spectrometer tube has been found to operate well at pressures from 10^{-6} to 10^{-4} mm. It should be built in a form which can be baked because, otherwise, the gases usually found in unbaked tubes can give strong background mass components in the observations. On the other hand, masses can be resolved at pressures up to values of the pressure where the mean free path of the ions is about equal to the length of the tube.

It is important to note that no slits are required. All the ions produced in the ion source are accelerated through the essentially open structure of the tube toward the stopping electrode. The current at the top of the principal peak in Fig. 12 corresponds to about five percent of all the ions produced in the ion source.

Experiments are continuing to determine whether the harmonic peaks shown in Fig. 12 can better be eliminated by some other selection of integer combination for the number of cycles between stages or by increasing the ratio of the grid distances to the hole-sizes in the grids. A number of simplifications of the tube and equipment are now being tested to determine which can be introduced without losing too much resolution for analytical applications.

Mr. Harry Diamond and Dr. Robert D. Huntoon, of this Bureau, have contributed much in encouragement in this development.

Strain Induced Grain Boundary Migration in High Purity Aluminum

PAUL A. BECK AND PHILIP R. SPERRY*
Department of Metallurgy, University of Notre Dame, Notre Dame, Indiana
(Received August 8, 1949)

A method has been developed for the study of grain boundary migration in high purity aluminum, which allows positive determination of the *direction* of grain boundary movements. By means of this method it was confirmed that the migration of grain boundaries resulting from surface energy takes place in the direction toward the centers of curvature of the boundaries. On the other hand, strain induced grain boundary migration, such as occurs in recrystallization, proceeds in a direction away from the centers of curvature of the moving boundaries. It was also found that certain strained grains serve as nuclei for strain-free grains of the same orientation which grow at the expense of neighboring strained grains. Most of the strain-free grains formed during the annealing of high purity aluminum cold rolled up to about 40 percent are produced by this mechanism of strain-induced boundary migration, without the formation of new nuclei.

RECENT investigations of Harker and Parker,[1] C. S. Smith,[2] and J. E. Burke[3] revealed some important aspects of the type of boundary migration connected with gradual grain growth. A distinctive feature of this type of migration is that the boundaries move *in the direction of their centers of curvature.* A typical example is shown in Fig. 1, which represents two successive stages in the migration of grain boundaries in an annealed high purity aluminum specimen. By means of a technique developed at this Laboratory,[4] the two successive stages are shown superimposed on each other. The driving energy of grain growth is the excess free energy associated with the grain boundary surfaces. Grain growth tends to decrease this free energy, thus allowing the structure to approach equilibrium, by decreasing the grain boundary curvatures and the total grain boundary surface per unit volume.

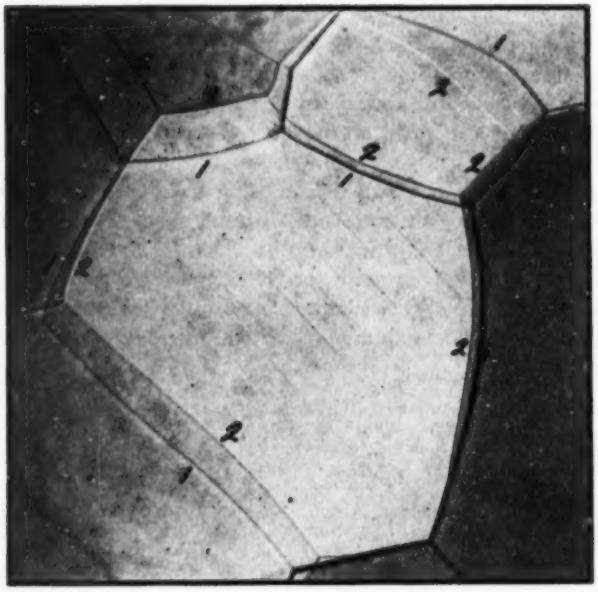


FIG. 1. In grain growth, boundary surfaces migrate toward their centers of curvature. Boundaries in high purity aluminum, after annealing 2 min. at 600°C (1), and after an additional anneal of 30 sec. at 600°C (2). Magnification 75×.

The recrystallization of cold worked metals also proceeds by means of boundary migration. However, in this case the moving boundaries separate strain hardened grains from annealed grains, and the direction of the movement is such that the volume of the annealed material increases with time at the expense of the cold worked material. The familiar concept of recrystallization involves the formation of nuclei of new strain-free grains in the midst of the strain hardened material. Most of the small new grains are at first entirely surrounded by one or two grains of the cold worked structure. During the growth of these new grains, their convex boundaries move in a direction *opposite to their centers of curvature.* Two stages of the growth of such a new grain (*C*) are shown in Fig. 2 by means of the same technique used in Fig. 1. The driving force of recrystallization is the excess free energy stored in the strain hardened material as a result of cold working. Even though the grain boundary movements here, as far as they can be observed under the microscope, take place in a direction opposite to that required by surface energy considerations, the total free energy is decreasing, and equilibrium is approached, as the strain hardened grains of high free energy content are replaced by low energy annealed grains.

Figure 2 also shows four successive stages of the migration of a grain boundary (1) already present in the original strain hardened structure. As revealed by the color effect in polarized light after electrolytic oxidation[4] and by x-ray diffraction, the orientation of the material swept by the moving boundary is very closely the same as that of the adjacent strain hardened grain (*A*). At first glance it might appear as if this boundary migration was related to ordinary grain growth, since one of the grains already present in the structure (*A*), is growing at the expense of another one

* Professor of Metallurgy, and Metallographer, respectively.
[1] D. Harker and E. R. Parker, "Grain shape and grain growth," Trans. A.S.M.E. **34**, 156 (1945).
[2] C. S. Smith, "Grains, phases, and interfaces," Trans. A.I.M.E. **175**, 15 (1948).
[3] J. E. Burke, "Some factors affecting the rate of grain growth in metals," Trans. A.I.M.E. **180**, 73 (1949).
[4] Philip R. Sperry, Trans. A.I.M.E. **188**, 103 (1949).

(*B*). On closer inspection, however, it becomes obvious that this process is different from grain growth, and that it is in important respects similar to recrystallization. This type of boundary migration takes place only in strain hardened material and it leads to an increase, rather than to a decrease, of grain boundary surface area, as seen in Fig. 2.

The question arises whether or not the observed boundary movements are motivated by the excess free energy of strain hardening, as in recrystallization. Figure 3 shows an example of the type of strain induced grain boundary migration discussed. The material swept by the moving boundary consists of a number of sections with very slightly different orientation. All these orientations are almost identical with the average orientation of the adjacent strain hardened grain (*A*). With the type of specimen preparation used, the surface of strain hardened grains becomes roughened, while the surface of strain free grains remains quite smooth. Figure 4 shows, by means of oblique illumination, that grains (*A*), (*B*) and (*C*), remaining from the original cold worked structure, have rough surfaces. On the other hand, the area swept by the moving boundary while migrating from position 1 to position 2, is quite smooth. As confirmed by means of x-ray diffraction, the material in this area is essentially in the annealed, strain-free, condition. It thus appears that in the strain induced boundary migration considered, as in ordinary recrystallization, the moving boundaries separate annealed material from strain hardened material and the direction of the boundary movement is such as to increase the volume of the annealed material. It is clear that the driving energy of the two processes is the same.

There is one significant difference, however, between the type of strain induced boundary migration shown

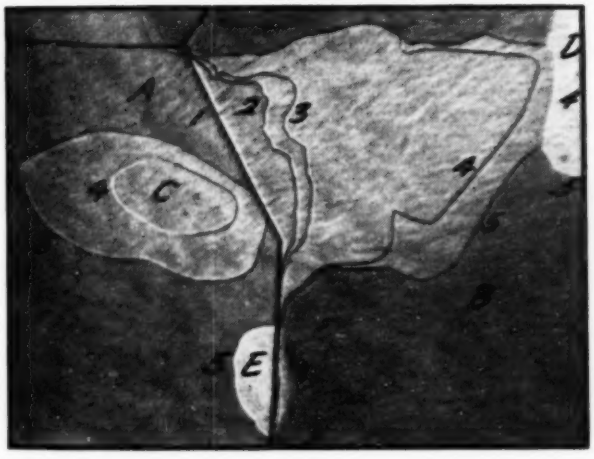


FIG. 2. In recrystallization (grains *C*, *D* and *E*) boundary surfaces migrate away from their centers of curvature. Strain induced growth of grain *A* at the expense of grain *B*. High purity aluminum annealed 90 min. at 600°C, rolled 7.5 percent (heavily outlined straight boundaries (1)), annealed successively 5 sec. (2), 5 sec. (3), 15 sec. (4), and 7 sec. (5) at 500°C. Magnification 75×.

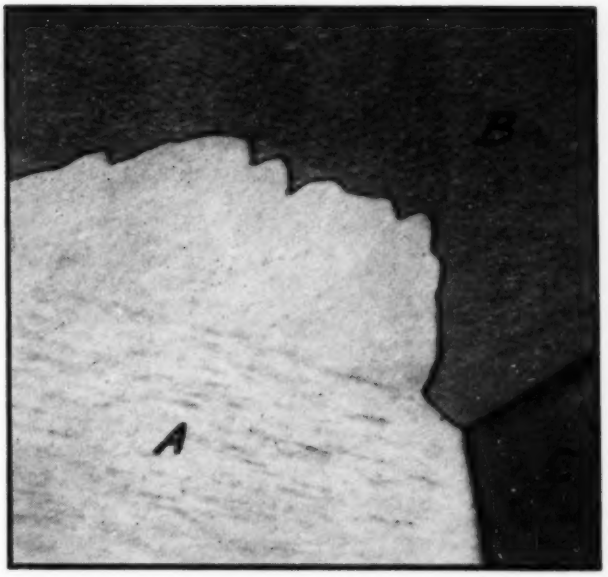


FIG. 3. Strain induced boundary migration in high purity aluminum. Orientation in area swept by migrating boundary is almost identical with that of strain hardened parent grain (A). Penultimate anneal: 5 min. at 650°C. 14 percent reduction by rolling. Last anneal 5 min. at 400°C. Magnification 75×.

in Fig. 3 and ordinary recrystallization. In the former, the annealed material swept by the migrating boundary has the same orientation as its strain hardened parent (*A* in Fig. 4), while in recrystallization a nucleus of different orientation is formed. As seen in Fig. 4, strain induced boundary migration without strain induced nucleation results in annealed areas bounded by an ordinary grain boundary on one side, and bordering on a strain hardened grain of the same orientation, without any intervening grain boundary, on the other side. That this is not so in ordinary recrystallization is shown in Fig. 5, where recrystallized grains *A* and *B* are located between strain hardened grains *C* and *D* and are separated from both of them by ordinary grain boundaries.

In summary, it may be stated that the processes of strain induced boundary migration found in aluminum is very much akin to recrystallization but that it differs from the latter by the fact that it is apparently nucleated by a strain hardened grain already present in the structure. Thus, *strain induced boundary migration may be considered as recrystallization without the formation of new nuclei*. Correspondingly, it was observed in many instances that, at the boundaries where it does occur, strain induced boundary migration starts promptly at the beginning of the annealing treatment. On the other hand, it is well known that a definite period of incubation is required for the formation of recrystallization nuclei and for their growth to microscopic sizes. This point is illustrated by Fig. 2, where each successive period of annealing has brought about a clearly observable migration of the boundary initially

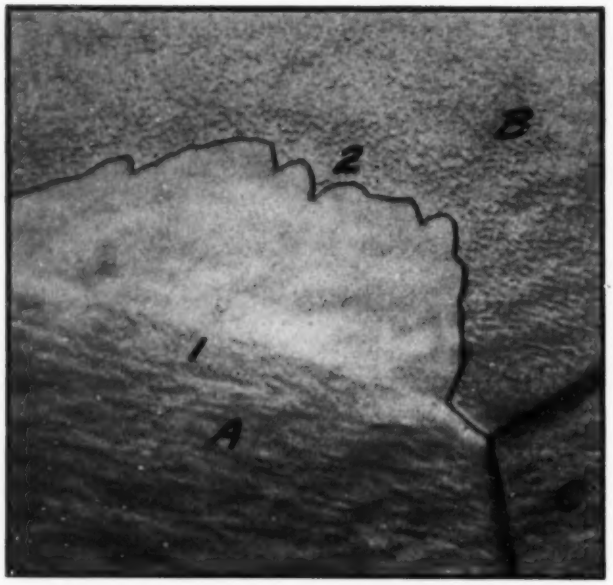


FIG. 4. Strain induced boundary migration in high purity aluminum, as in Fig. 3. Strain hardened grains (*A*), (*B*), and (*C*) are roughened by etching, strain-free area, swept by migrating boundary, is smooth, and has no grain boundary toward grain (*A*). Oblique illumination.

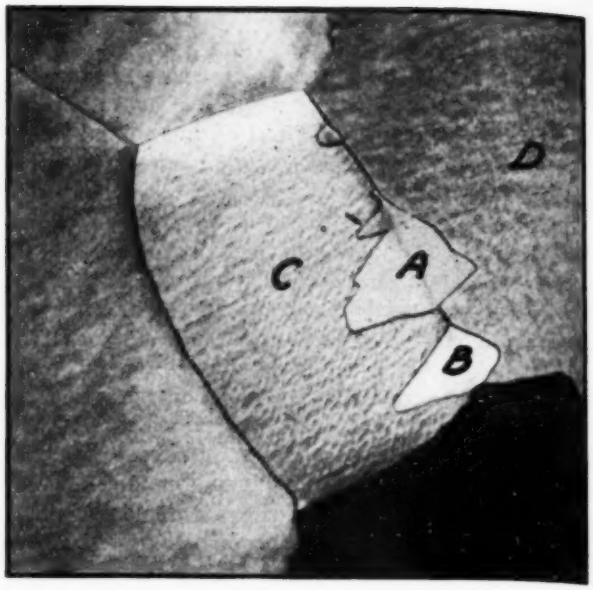


FIG. 5. Recrystallized grains (*A*) and (*B*) are smooth and have grain boundaries toward both grains (*C*) and (*D*). The latter are strain hardened and consequently roughened by etching. Magnification 75×.

situated at (1). However, a recrystallized grain (*C*) came into existence only during the third period of annealing, and its two boundary positions, seen in the picture, correspond to the end of the last two annealing periods. Similarly, another recrystallized grain (*E*) was born during the fourth period of annealing and, therefore, the figure shows only one boundary position (5) for this grain. It was found that the ordinary recrystallization process, involving nucleation, plays a relatively minor role during the annealing of high purity aluminum deformed by rolling up to about 40 percent reduction of area. In such instances the type of strain induced boundary migration not involving the formation of new nuclei appears to predominate in the transformation to the strain-free condition.

This work was supported by the ONR, U. S. Navy, Contract No: N6 ori-165, T.O. 1.

The External Field Produced by a Slot in an Infinite Circular Cylinder*

SAMUEL SILVER AND WILLIAM K. SAUNDERS
Division of Electrical Engineering and the Antenna Laboratory, University of California, Berkeley, California
(Received August 10, 1949)

Expressions are derived for the external field produced by a slot of arbitrary shape in the wall of a circular wave guide (of infinite extent and infinite conductivity), the tangential components of the electric field in the slot being assumed to have been prescribed. This is accomplished by matching a Fourier representation of the external field, built up by superposition of basic sets of cylindrical waves, to a Fourier expansion of the prescribed field over the surface of the cylindrical wave guide. The far-zone field is obtained by applying the method of steepest descent to the Fourier integrals that constitute the coefficients in the expansion for the external field. The results satisfy the radiation conditions for far-zone fields.

I. INTRODUCTION

THIS paper, the first of a series of studies of the radiative properties of slots in conducting surfaces, is concerned with the field set up in space by a slot of arbitrary shape in the wall of a circular cylindrical wave guide. A complete analysis of the properties of a slot in the wall of a wave guide must make proper connection between the interior and exterior fields, thus determining the field distribution in the slot itself. Such a unified treatment is difficult to carry out and as yet has not been achieved for even the simplest slot configurations. Instead, the procedure has been to treat the interior and exterior problems separately in terms of assumed field distributions in the slot,[1] a procedure that we follow here. Under certain conditions, as for example in the case of an elongated narrow slot, reasonable assumptions can be made concerning the field distribution in the slot and results can be obtained that are of considerable usefulness in the design of radiating systems.

The subject of the slot in a circular wave guide has been treated previously under various restrictive conditions on its geometry and field distribution. G. Sinclair[2] and C. H. Papas and R. W. P. King[3] have discussed the field of a narrow slot of infinite length running parallel to the axis of the cylinder, excited uniformly along its length in the transverse direction. More recently C. H. Papas,[4] by means of a Green's function technique, has solved the problem of a finite slot in which the electric field has but a single tangential component that is parallel to the axis of the cylinder. This essentially restricts the configuration of the slot to be developable into a rectangle oriented with its boundaries respectively parallel to and perpendicular to the axis of the cylinder. As we shall show in Section V, the several previous results can be obtained directly from our general expressions.

II. GENERAL CONSIDERATIONS

The analysis presented here is free from the restrictive conditions on the slot geometry and fields that characterize the earlier work but is based also on the assumptions that the wave guide is infinite in extent and has infinite conductivity. The time dependence of the fields is taken to be harmonic and the complex exponential $e^{+j\omega t}$ is chosen as its representation.

The axis of the cylinder, which has a radius a, serves as the z axis of our coordinate system. Cylindrical coordinates ρ, ϕ, z are most suited for developing the general solution. Later, for purposes of describing the far-zone field, spherical coordinates R, θ, ϕ will be more appropriate; the cylinder axis will then form the polar axis of the spherical coordinate system. On the surface of the cylinder we have a slot of arbitrary shape, bounded in the axial direction by the planes $z=z_1$ and $z=z_2$ and defined circumferentially by the curves $\phi_1(z)$ and $\phi_2(z)$; the various quantities are shown in Fig. 1.

The tangential electric field in the slot will in general have both ϕ and z components which we consider to be prescribed functions $f_1(\phi, z)$ and $f_2(\phi, z)$, respectively. In view of the infinite conductivity of the wall

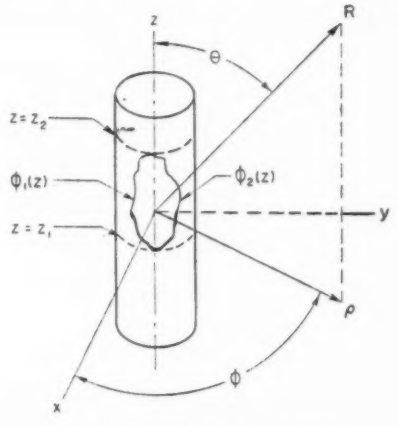


FIG. 1. Geometrical parameters.

* This work was supported in part by the Bureau of Ships, Contract NObsr-39401.

[1] For a general development of the interior problem see A. F. Stevenson, J. App. Phys. **19**, 24 (1948).

[2] G. Sinclair, Proc. I.R.E. **36**, 1487 (1948).

[3] C. H. Papas and R. W. P. King, Cruft Laboratory Technical Report No. 32, ONR Contract N5-ori-76 (March, 1948); Quart. App. Math. **7**, No. 2 (July, 1949).

[4] C. H. Papas, Cruft Laboratory Technical Report No. 58, ONR Contract N5-ori-76 (September, 1948); J. App. Phys. **20**, 437 (1949); Cruft Laboratory Technical Report No. 61, ONR Contract N5-ori-76 (October, 1948) to be published in J. Math. and Phys. **28**, No. 4 (January, 1950).

this represents a complete knowledge of the tangential electric field over the entire surface of the cylinder, namely,

$$E_\phi = f_1(\phi, z) \quad \phi_1 \leq \phi \leq \phi_2;\ z_1 \leq z \leq z_2;$$
$$= 0 \qquad \text{outside the slot;} \tag{1a}$$

$$E_z = f_2(\phi, z) \quad \phi_1 \leq \phi \leq \phi_2;\ z_1 \leq z \leq z_2;$$
$$= 0 \qquad \text{outside the slot.} \tag{1b}$$

A solution of Maxwell's equations must be constructed that assumes the prescribed value of the tangential electric field over the cylinder and that satisfies certain radiation conditions at infinity. For a time-periodic field the latter take the form:[5]

$$\lim_{R\to\infty} R\mathbf{E} \text{ is finite,} \tag{2a}$$

$$\lim_{R\to\infty} R[\mathbf{R}_1 \times \mathbf{H} + (\epsilon/\mu)^{\frac{1}{2}}\mathbf{E}] = 0, \tag{2b}$$

$$\lim_{R\to\infty} R\mathbf{H} \text{ is finite,} \tag{2c}$$

$$\lim_{R\to\infty} R[(\mathbf{R}_1 \times \mathbf{E}) - (\mu/\epsilon)^{\frac{1}{2}}\mathbf{H}] = 0, \tag{2d}$$

where $\mathbf{R}_1$ is a radial unit vector in the spherical coordinate system (see Fig. 1). Equations (2b) and (2d) are the radiation conditions proper and express the fact that we have only outgoing waves at infinity. A time-periodic solution that meets the above requirements is unique.

The procedure is to construct Fourier representations for the tangential electric field over the cylinder and for the field in space. The latter is synthesized by superposition of basic sets of cylindrical waves that satisfy the requisite conditions at infinity. The coefficients of the expansion for the surface field are known; the coefficients of the representation for the space field are determined by the requirement that its tangential components at the surface of the cylinder shall reduce to the Fourier representation of the field prescribed over the latter. The resulting expressions are in the form of a Fourier series whose coefficients are Fourier integrals.

III. DEVELOPMENT OF THE FIELDS

The Tangential Field over the Cylinder

We shall first develop the Fourier expansion for the tangential components of the electric field over the surface of the cylinder. Let f_α denote either $f_1(\phi, z)$ or $f_2(\phi, z)$ and let E_α be the corresponding field component. In the ϕ-direction E_α is a periodic function and is therefore to be represented by a Fourier series

$$E_\alpha(\phi, z) = \sum_{n=-\infty}^{\infty} A_n{}^\alpha(z) e^{-jn\phi}, \tag{3}$$

the coefficients $A_n{}^\alpha$ being functions of z. It is readily evident that

$$A_n{}^\alpha = \frac{1}{2\pi} \int_{\phi_1(z)}^{\phi_2(z)} f_\alpha(\beta, z) e^{jn\beta} d\beta, \quad z_1 \leq z \leq z_2; \tag{4a}$$

$$= 0. \qquad z < z_1, \quad z > z_2. \tag{4b}$$

$A_n{}^\alpha(z)$ is thus a piecewise continuous function and its Fourier representation is the Fourier integral

$$A_n{}^\alpha = \frac{1}{2\pi} \int_{-\infty}^{\infty} dh \int_{z_1}^{z_2} A_n{}^\alpha(\xi) e^{-jh(z-\xi)} d\xi; \tag{5}$$

or,

$$A_n{}^\alpha(z) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dh \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_\alpha(\beta, \xi) e^{jn\beta} e^{-jh(z-\xi)} d\beta. \tag{6}$$

The tangential electric field over the cylinder is thus expressed in the form of a Fourier expansion:

$$E_\alpha(\phi, z) = \frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} e^{-jn\phi} \int_{-\infty}^{\infty} dh \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_\alpha(\beta, \xi) e^{jn\beta} e^{-jh(z-\xi)} d\beta. \tag{7}$$

The External Field Expansion

We construct the external field by superposition of basic sets of cylindrical waves. The reader is referred to Stratton[6] for an exposition of the general theory of the latter. The basic sets fall into two groups: the transverse magnetic (TM) type having no H_z component and the transverse electric (TE) type having no E_z component. Appropriate combinations of these can be made to form a basic set of field components from which a general field that is periodic in ϕ can be synthesized. Each member of the latter set is a Fourier series as follows:

$$\mathcal{E}_\rho = \left\{ \sum_{n=-\infty}^{\infty} \left[-jha_n \frac{\partial H_n{}^{(2)}(\Lambda\rho)}{\partial\rho} - \frac{n\omega\mu}{\rho} b_n H_n{}^{(2)}(\Lambda\rho) \right] e^{-jn\phi} \right\} e^{-jhz}, \tag{8a}$$

$$\mathcal{E}_\phi = \left\{ \sum_{n=-\infty}^{\infty} \left[-\frac{nh}{\rho} a_n H_n{}^{(2)}(\Lambda\rho) + j\omega\mu b_n \frac{\partial H_n{}^{(2)}(\Lambda\rho)}{\partial\rho} \right] e^{-jn\phi} \right\} e^{-jhz}, \tag{8b}$$

[5] S. Silver, *Microwave Antenna Theory and Design* (McGraw-Hill Book Company, Inc., New York, 1949), Radiation Laboratory Series, Vol. 12, Chapter 3.

[6] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), Chapter 6. Note that our expressions differ from those of Stratton in signs because of our choice of $e^{+j\omega t}$ rather than $e^{-j\omega t}$ to represent the time dependence. This also leads to the use of the Hankel function of the second kind to represent an outward traveling wave.

$$\mathcal{E}_z=\left\{\sum_{n=-\infty}^{\infty}[(k^2-h^2)a_nH_n^{(2)}(\Lambda\rho)]e^{-jn\phi}\right\}e^{-jhz}, \quad (8c)$$

$$\mathcal{H}_\rho=\left\{\sum_{n=-\infty}^{\infty}\left[\frac{nk^2}{\mu\omega\rho}a_nH_n^{(2)}(\Lambda\rho)-jhb_n\frac{\partial H_n^{(2)}(\Lambda\rho)}{\partial\rho}\right]e^{-jn\phi}\right\}e^{-jhz}, \quad (8d)$$

$$\mathcal{H}_\phi=\left\{\sum_{n=-\infty}^{\infty}\left[-j\frac{k^2}{\mu\omega}a_n\frac{\partial H_n^{(2)}(\Lambda\rho)}{\partial\rho}-\frac{nh}{\rho}b_nH_n^{(2)}(\Lambda\rho)\right]e^{-jn\phi}\right\}e^{-jhz}, \quad (8e)$$

$$\mathcal{H}_z=\left\{\sum_{n=-\infty}^{\infty}[(k^2-h^2)b_nH_n^{(2)}(\Lambda\rho)]e^{-jn\phi}\right\}e^{-jhz}, \quad (8f)$$

with

$$\Lambda=[k^2-h^2]^{\frac{1}{2}}. \quad (9)$$

In the case of harmonic time dependence k is a constant that is given in terms of the constitutive parameters—magnetic permeability μ, electric permittivity ϵ, conductivity σ—of the medium under consideration and the angular frequency by

$$k=[\omega^2\mu\epsilon-j\mu\sigma\omega]^{\frac{1}{2}}; \quad -\pi/2<ph\,k\leq 0. \quad (10)$$

The parameter h with respect to which the set is ordered is in general complex; its values are to be taken to suit the required boundary conditions. The Hankel function of the second kind enters into the wave functions to represent outgoing waves. However, this property of the Hankel functions depends on the definition of the phase of Λ for the various values of h. We shall return to this matter shortly.

The terms in Eqs. (8) having the common coefficients a_n belong to the set of TM waves; those having common coefficients b_n are of the TE type. It will be observed that there are only two sets of independent coefficients and hence, a knowledge of any two field components is sufficient to establish the complete field. The coefficients a_n and b_n are independent of ρ and z but in general are functions of h.

We can now build up an arbitrary z dependence by superposition of components of various values of h. Regarding h as a complex quantity, different representations can be obtained according to the path in the h plane along which h is allowed to vary. In the present case where we are to fit the field at $\rho=a$ to a Fourier integral along the surface of the cylinder the path along the real-h axis is indicated. Thus for any field component $E_\alpha(\rho, \phi, z)$ we shall write

$$E_\alpha(\rho, \phi, z)=\int_{-\infty}^{\infty}\mathcal{E}_\alpha(\rho, \phi, h)dh. \quad (11)$$

Similar expressions hold for the magnetic field components.

We have noted before that the phase of $\Lambda=(k^2-h^2)^{\frac{1}{2}}$ must be defined along the path of integration in such a way as to produce proper behavior of the field at infinity. The locations of the branch points $h=\pm k$ are shown in Figs. 2(a) and 2(b), the latter applying to the case when the external medium has neither an ohmic nor a dielectric dissipation factor. To define the phase of Λ let us consider the behavior of $H_n^{(2)}(\Lambda\rho)$ for very large values of ρ, making use of the asymptotic form of the Hankel function,

$$H_n^{(2)}(\Lambda\rho)\approx[2/\pi\Lambda\rho]^{\frac{1}{2}}e^{-j\Lambda\rho}e^{j(2n+1)\pi/4}. \quad (12)$$

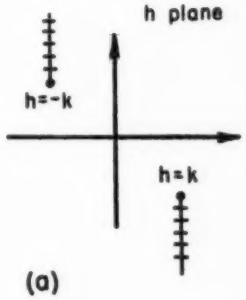
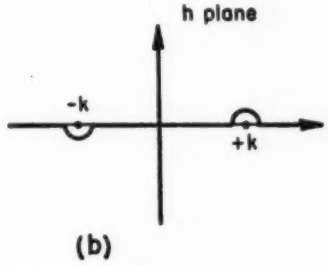


FIG. 2. Path of integration in the h Plane: (a) k complex; (b) k real.



For the general case, when k is complex, it is seen that it is necessary to require

$$Re\Lambda\geq 0, \quad Im\Lambda\leq 0 \quad (13)$$

in order to have outward radial propagation and to keep the field components bounded at infinity. If k is real, the conditions (13) reduce to

$$ph\Lambda=0 \qquad |h|<k, \quad (14a)$$

$$ph\Lambda=-\pi/2 \qquad |h|>k, \quad (14b)$$

and it is seen that the contour must be indented around the branch points in the manner shown in Fig. 2(b).

Evaluation of the Coefficients

With the sense of Λ properly defined the components of the external field are given by Eq. (11). To evaluate

the coefficients a_n and b_n we match the components E_ϕ and E_z to the assigned values over the cylinder. From Eqs. (8) and (11), on interchanging summation and integration, we obtain

$$E_\phi = \sum_{n=-\infty}^{\infty} e^{-jn\phi} \int_{-\infty}^{\infty} \left[-\frac{nh}{\rho} a_n H_n^{(2)}(\Lambda\rho) + j\omega\mu b_n \frac{\partial H_n^{(2)}(\Lambda\rho)}{\partial\rho} \right] e^{-jhz} dh, \quad (15a)$$

$$E_z = \sum_{n=-\infty}^{\infty} e^{-jn\phi} \int_{-\infty}^{\infty} (k^2 - h^2) a_n H_n^{(2)}(\Lambda\rho) e^{-jhz} dh. \quad (15b)$$

At $\rho = a$ these must reduce to the expansions of Eq. (7); hence,

$$\left[-\frac{nh}{a} H_n^{(2)}(\Lambda a) \right] a_n + \left[j\omega\mu \frac{\partial H_n^{(2)}(\Lambda a)}{\partial a} \right] b_n = \frac{1}{4\pi^2} \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_1(\beta, \xi) e^{jn\beta} e^{jh\xi} d\beta, \quad (16a)$$

$$[(k^2 - h^2) H_n^{(2)}(\Lambda a)] a_n = \frac{1}{4\pi^2} \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_2(\beta, \xi) e^{jn\beta} e^{jh\xi} d\beta; \quad (16b)$$

or,

$$a_n = \frac{1}{4\pi^2 (k^2 - h^2) H_n^{(2)}(\Lambda a)} \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(}^{\phi_2(\xi)} f_2(\beta, \xi) e^{jn\beta} e^{jh\xi} d\beta, \quad (17a)$$

$$b_n = \frac{-jnh}{4\pi^2 \omega\mu a (k^2 - h^2) (\partial H_n^{(2)}(\Lambda a)/\partial a)} \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_2(\beta, \xi) e^{jn\beta} e^{jh\xi} d\beta - \frac{j}{4\pi^2 \omega\mu (\partial H_n^{(2)}(\Lambda a)/\partial a)} \int_{z_1}^{z_2} d\xi \times \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_1(\beta, \xi) e^{jn\beta} e^{jh\xi} d\beta. \quad (17b)$$

These taken together with Eqs. (8) and (11) give the formal solution to the problem. The surface current distribution **K** over the cylinder is given directly by the tangential magnetic field at the surface:

$$K_\phi = -H_z(a, \phi, z); \quad K_z = H_\phi(a, \phi, z). \quad (18)$$

Whereas the process of obtaining the formal solution is a relatively simple one, the numerical evaluation of the fields is a task of considerable magnitude, particularly for points in the neighborhood of the slot. However, for the far-zone field the expressions simplify to a large extent and numerical work becomes more tractable. We shall discuss this region of the field in the following section.

IV. THE FAR-ZONE FIELD

The reduction of the general expressions for this region is effected by making use of the asymptotic forms of the Hankel functions and subsequently carrying out certain integrations by the saddle-point method.[8] We shall restrict the treatment to the case when k is real.

It is convenient to describe the far-zone field in terms of spherical coordinates R, θ, ϕ (see Fig. 1), and, accordingly we make the substitutions

$$\rho = R\sin\theta, \quad z = R\cos\theta \quad (19)$$

in the expressions for the components of the field. We are interested in the dominant characteristics of the field at large distances R; that is, we wish to obtain the leading terms in a development of the field components in inverse powers of R. Except for small values of θ we may, for large values of R, replace $H_n^{(2)}(\Lambda\rho) = H_n^{(2)}(\Lambda R\sin\theta)$ by the asymptotic form given in Eq. (12).† When this is done and the order of integration is suitably rearranged, it is found that the expressions for the field components take the general form

$$\frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty}{}' e^{-jn\phi} \left\{ \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_1(\beta, \xi) e^{jn\beta} d\beta \times \int_{-\infty}^{\infty} u(h, R, \theta; n) e^{jh\xi} e^{-jR[\Lambda \sin\theta + h\cos\theta]} dh + \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} f_2(\beta, \xi) e^{jn\beta} d\beta \times \int_{-\infty}^{\infty} v(h, R, \theta; n) e^{jh\xi} e^{-jR[\Lambda \sin\theta + h\cos\theta]} dh \right\}. \quad (20)$$

The functions u and v can readily be evaluated for each particular component. It will be observed that they vary slowly over the range of h as compared with the exponential functions. The integrals over h in (20) are, therefore, of the general type to which the saddle-point method is applicable. We are here primarily concerned with the region of the field where R is very

[8] See, for example, G. N. Watson, *Bessel Functions* (Cambridge University Press, London, 1944), pp. 235 ff.

† We shall not enter here into a consideration of the different asymptotic representations that must be used according as to whether the order of the Hankel function is small, nearly equal to, or large compared with the magnitude of $\Lambda\rho$. From a practical standpoint these details are not too important inasmuch as the terms of the series are found to diminish rapidly with increasing values of n.

much larger than both the wave-length and the axial extent of the slot ($R \gg \xi$). The variation of the phase factor involving R thus dominates that involving ξ and we have the simplest form of saddle-point problem. It will be well to note in passing that in the case of a slot whose axial extent is large compared with the wave-length there is an intermediate zone, at distances about equal to or possibly less than the total axial dimension, for which (20) also gives the dominant terms. In the latter case, however, we have two rapidly varying phase factors to consider in the integration over h and consequently a saddle-point problem involving two large parameters.‡ Inasmuch as the intermediate zone

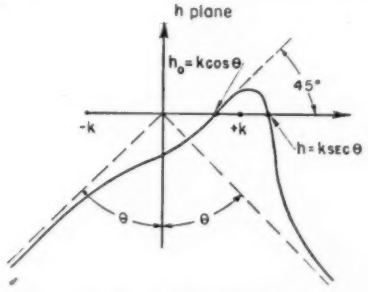


FIG. 3. Modified path through the saddle-point.

is of limited interest it will not be discussed further in this paper. Practical interest usually centers about the far-zone field pattern which is one of the primary specifications that are laid down for the design of a given antenna system.

With the restriction $R \gg \xi$ we may formulate our problem as that of evaluating an integral of the form

$$\int_{-\infty}^{\infty} F(h)e^{-jR[\Lambda \sin\theta + h\cos\theta]}dh = \int_{-\infty}^{\infty} F(h)e^{R\Phi(h)}dh \quad (21)$$

for large values of R, with

$$F(h) = \left.\begin{matrix} u(h, R, \theta; n) \\ v(h, R, \theta; n) \end{matrix}\right\} e^{jh\xi}. \quad (21a)$$

The saddle-point technique consists of deforming the path of integration into one that passes through the point h_0 for which $\Phi'(h_0)=0$ and is such that over its entirety

$$Im\Phi(h) = Im\Phi(h_0).$$

If the integral converges along this path, the major contribution to the integral arises from a small segment about the saddle-point $h=h_0$. In our case the new path is a function of θ; its general form is shown in Fig. 3. The deformation from the real axis to this path is possible by virtue of the analytic properties of $F(h)$ and by the extension of conditions such as were imposed on Λ in Eq. (13) to the entire exponent, namely $Im\Phi(h)<0$. This makes the integral over the arc connecting the real axis and the new path at infinity vanish, and also ensures the convergence of the integral along the modified path. Considering the small segment of the latter in the neighborhood of h_0 having the modulus $u<\epsilon$, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} F(h)e^{R\Phi(h)}dh \\
&\quad \approx F(h_0)e^{[R\Phi(h_0)+j\pi/4]} \int_{-\epsilon}^{\epsilon} e^{j\Phi''(h_0)u^2/2}du \\
&\quad = F(k\cos\theta)e^{-j(kR-\pi/4)} \int_{-\epsilon}^{\epsilon} e^{-Ru^2/2k\sin^2\theta}du \\
&\quad \approx F(k\cos\theta)e^{-j(kR-\pi/4)}\left[\frac{2k\sin^2\theta}{R}\right]^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\tau^2}d\tau;
\end{aligned}$$

or,

$$\int_{-\infty}^{\infty} F(h)e^{R\Phi(h)}dh \approx \left[\frac{2\pi k}{R}\right]^{\frac{1}{2}} \sin\theta F(k\cos\theta)e^{-jkR}e^{j\pi/4}. \quad (22)$$

The leading terms are found to be of order R^{-1}. These form the radiation field. The components of E_ρ and E_z in the R-direction that are of this order of magnitude cancel one another; the same occurs in the magnetic field. The radiation field is thus transverse to the R-direction and, in general, has non-zero components:

$$E_\theta = -\frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\sum_{n=-\infty}^{\infty}\frac{j^{n+1}e^{-jn\phi}}{\sin\theta H_n^{(2)}(ka\sin\theta)} \times \int_{z_1}^{z_2} d\xi e^{jk\xi\cos\theta}\int_{\phi_1(\xi)}^{\phi_2(\xi)} f_2(\beta,\xi)e^{jn\beta}d\beta; \quad (23a)$$

$$\begin{aligned}
E_\phi = \frac{1}{2\pi^2}\frac{e^{-jkR}}{R}\sum_{n=-\infty}^{\infty}\frac{j^n e^{-jn\phi}}{H_n^{(2)\prime}(ka\sin\theta)} &\times\left[\int_{z_1}^{z_2} d\xi e^{jk\xi\cos\theta}\int_{\phi_1(\xi)}^{\phi_2(\xi)} f_1(\beta,\xi)e^{jn\beta}d\beta\right. \\
&+\frac{n\cot\theta}{ka\sin\theta}\int_{z_1}^{z_2} d\xi e^{jk\xi\cos\theta} \\
&\left.\times\int_{\phi_1(\xi)}^{\phi_2(\xi)} f_2(\beta,\xi)e^{jn\beta}d\beta\right]; \quad (23b)
\end{aligned}$$

$$H_\theta = -(\epsilon/\mu)^{\frac{1}{2}}E_\phi; \quad H_\phi = (\epsilon/\mu)^{\frac{1}{2}}E_\theta. \quad (23c)$$

These show explicitly that the solutions satisfy the required conditions at infinity that are formulated in Eqs. (2).

‡ The problem is very similar to that of obtaining various asymptotic forms of the cylinder functions according to the relative magnitudes of the order and the argument of the functions; see reference (8), Chapter 8.

V. ILLUSTRATIVE EXAMPLES

To illustrate the applicability of the results we shall touch briefly upon two cases reported previously in the literature to show how the earlier results are readily obtained from our general expressions:

1. The Uniformly Excited Circumferential Slot[4]

The slot runs completely around the cylinder and has a height $2w$ that is very much smaller than the circumference. For convenience we shall take $z_1=-w$, $z_2=w$. Note also that $\phi_1(z)=0$, $\phi_2(z)=2\pi$. The tangential electric field in the slot has only a z component and that independent of ϕ; thus

$$f_1(\phi, z)=0; \quad f_2(\phi, z)=E_0. \tag{24}$$

The integrals over β in Eqs. (17) vanish except for $n=0$. Consequently, the only non-zero coefficient is

$$a_0=\frac{2E_0w(\sin hw/hw)}{2\pi(k^2-h^2)H_0^{(2)}(\Lambda a)}. \tag{25}$$

Defining the voltage across the slot to be $V=2E_0w$, we get for the general expression for the z component of the field in space

$$E_z(\rho, \phi, z)=\frac{V}{2\pi}\int_{-\infty}^{\infty}\frac{H_0^{(2)}[(k^2-h^2)^{\frac{1}{2}}\rho]}{H_0^{(2)}[(k^2-h^2)^{\frac{1}{2}}a]}\frac{\sin hw}{hw}e^{-jhz}dh. \tag{26}$$

If the slot is a δ-slot, that is, $w\to 0$ while V is constant, $\sin hw/hw$ reduces to unity. The far-zone field is also readily obtained by evaluating the $n=0$ term of Eq. (23a) for the given $f_2(\phi, z)$.

2. The Infinite Axial Slot[2, 3]

Another case that has been dealt with in the literature is that of a narrow rectangular slot, parallel to the axis of the cylinder, over which the field has only an E_ϕ component. The width of the slot will be taken to be $2a\phi_0$. The assumed field distribution in the slot is

$$f_1(\phi, z)=E_0, \quad f_2(\phi, z)=0. \tag{27}$$

All the coefficients a_n are now equal to zero, while for b_n we get

$$b_n=\lim_{l\to\infty}\frac{E_0}{j4\pi^2\omega\mu(\partial H_n^{(2)}(\Lambda a)/\partial a)}\times\int_{-l}^{l}e^{j\lambda\xi}d\xi\int_{-\phi_0}^{\phi_0}e^{jn\beta}d\beta. \tag{28}$$

Defining a voltage $V=2E_0a\phi_0$ and making use of the Fourier integral definition of the δ-function,

$$\delta(h-0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\lambda\xi}d\xi \tag{29}$$

we obtain finally for the E_ϕ component of the field

$$E_\phi=\frac{V}{2\pi a}\sum_{n=-\infty}^{\infty}\frac{\sin n\phi_0}{n\phi_0}\frac{H_n'^{(2)}(k\rho)}{H_n'^{(2)}(ka)}e^{-jn\phi}. \tag{30}$$

If the slot is regarded as a δ-slot in the ϕ-direction, $\sin n\phi_0/n\phi_0\to 1$ for all n.

Further application of the theory will be presented in a forthcoming paper in which we discuss the transverse rectangular slot in detail and give experimental data in conjunction with the theoretical results.

The Reversal Theorem of Linearized Supersonic Airfoil Theory*

M. M. Munk
Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received May 27, 1949)

It is demonstrated on simple dynamic principles that a supersonic airfoil under certain conditions exhibits the same drag and the same lift slope regardless of whether it flies forward or backward.

THE thin airfoil theory in the supersonic regime is ordinarily referred to as the linearized theory. The boundary condition is simplified exactly as in the case of the incompressible fluid. In addition, the partial differential equation is also simplified, and it reduces to the wave equation:

$$\frac{\partial^2 F}{\partial x^2}+\frac{\partial^2 F}{\partial y^2}-\frac{\partial^2 F}{\partial z^2}=0. \tag{1}$$

Herein, F denotes either the velocity potential, any one Cartesian velocity component, or the pressure. The variables x, y, and z denote Cartesian coördinates in spanwise, liftwise, and fore-aft-wise direction of an airfoil, respectively. The corresponding components of the increment velocity are u, v, and w. The increment velocity is small when compared with the velocity of advance W_0.

In the computations of this theory, the density ρ_0 occurs as if it were constant. Absence of divergence of velocity is not assumed, not even approximately, in conjunction with such constant density. The increment pressure Δp is

$$\Delta p=-\rho_0 w W_0. \tag{2}$$

The rate at which energy is left behind the airfoil advancing with the velocity W_0 in air otherwise at rest is equal to DW_0, the product of the drag and of the velocity of advance. We assume specifically that the drag as defined from this relation is consistent with what follows, especially with the absence of a finite chordwise, leading-edge force. Since Eq. (1) is linear, the superposition of any number of individual solutions will again be a solution.

The vortex theorems hold. All vortices within the air are parallel to the Z direction. No finite force acting on an infinitely thin leading edge in the chordwise direction is supposed to exist. The opposite would be inconsistent with finite bounds for the pressure or suction. The resultant airfoil force is, therefore, merely the sum total of the ordinary pressure effects. With a constant local angle of attack of a plane, infinitely thin airfoil, the resultant airforce is at right angles to the airfoil. The constant angle of attack is equal to the drag-lift ratio. Both the leading edge and the trailing edge are points or lines of separation or rejoining of the flow. It is assumed that this double-edge condition is not inconsistent with the existence of a solution of Eq. (1). The solution is not made unique by selecting the trailing edge as a rejoining line. Rather, the solution is made unique by stipulating that all free vortices be trailing behind the airfoil and that all pressure waves proceed in the rear of the airfoil. In other words, all perturbations are restricted to the region downstream of the leading edge of the airfoil. With a large sweepback angle such as occurs with arrow wings, there may not exist any solution of (1) if both the leading separation line and the trailing separations line are arbitrarily specified. This is the case of the "subsonic" leading edge, where the downstream Mach cones, having their apices at the trailing edge, intersect with or enclose the leading edge. In that case it is necessary that the leading separation line be compatible with the other geometric conditions including the position of the trailing separation line, that is the trailing edge and with Eq. (1).

There exists an infinity of solutions that do not comply with this last stipulation. All of these solutions are acceptable on physical grounds and occasionally represent what happens or what may happen.

Reversing all velocities or all increment velocities leads again to a solution of (1).

Equation (1) is the basic equation in acoustics. Therefore, it has been stated that acoustical problems are analogous to the airfoil problems now considered. That is an extreme understatement: the linearized airfoil problem is itself an acoustical problem. The solutions exhibit all ordinary features of acoustical phenomena, wave effects, focusing effects, and that pronounced absence of harmony and of the continuity of analytic functions peculiar to acoustics. The solutions are disharmonic and non-unitary—a patchwork of heterogeneous pressure distributions of contrasting character. The computation has to be adapted to this state of things. One region after another region has to be computed in piecemeal fashion. No individual solution obtained may be applied to a slightly modified airfoil or Mach number by the application of the intuitive faculty along without careful scrutiny.

Notwithstanding their erratic character, the solutions are subordinated to simple, broad, and universal principles, transcending the irregularity in detail. Consider for instance a plane strip of constant width b extending streamwise into infinity. Its lift is finite, not even depending on the Mach number, such that the angle of

* Sponsored by the Office of Naval Research.

attack is equal to

$$\alpha = \frac{2L}{\rho_0 W_0^2 b^2 \pi}. \tag{3}$$

The spanwise lift distribution approaches (summed up streamwise) the elliptic lift distribution. The lift density asymptotically approaches zero in downstream direction, but it approaches it in a kind of periodic fashion and by no means monotonically.

If the strip is cambered with straight parallel span lines, the angle of attack at the leading edge is zero, the streamwise curvature also approaches zero and the angle of attack approaches asymptotically the above value (3). Then the drag will assume the minimum induced drag of the incompressible fluid theory. This is a very academic point, indeed, for the large airfoil area required would lead to excessive and dominating surface friction drag forces. Still, this illuminates the conditions and exemplifies the simple relations transcending the erratic details.

Many individual solutions of (1) relating to airfoils of various planviews and of wind sections have been determined. Those working in this field observed that all solutions are consistent with a remarkable symmetry principle. [See Bibliography.] Any airfoil moved under otherwise equal conditions forward and backward yields equal drag in these linearized computations. Turning the airfoil by 180° about a spanwise direction changes in general the lift distribution and the drag distributions, and sometimes very radically too. But the computed sum total of the drag remain always the same. With a plane, infinitely thin airfoil the computed lift also remains the same, so that for any airfoil the computed lift slope remains the same.

We proceed now to present a general proof of this reversal theorem based on very simple grounds. Picture the airfoil in place and expose it to the undisturbed flow (A). An increment flow (a) in rear of the airfoil is then produced. Leave the airfoil in place, but expose it now to an undisturbed airflow (B) having equal but opposite velocity as (A). This time an increment flow (b) will be created downstream of the airfoil in relation to (B), but upstream of the airfoil in relation to (A). The two increment flows (a) and (b) are ordinarily quite different, but not wholly uncorrelated. Combine or superpose now the flows (A), (a), and (b). This combination flow is now associated with free vortices and with pressure waves upstream of the airfoil as well as with free vortices and with pressure waves downstream of the airfoil. Right at the airfoil surface, all liftwise increment velocities at right angles to the airfoil cancel out. The airfoil consistent with the superposition flow is, therefore, plane, infinitely thin, and has zero incidence. The airfoil still has a lift distribution, but it cannot have any drag. The rate of energy transport upstream of the airfoil is equal to that of the flow (B)–(b). The rate of energy transport downstream of the airfoil is that of the flow (A)–(a). Both are alike, because the drag of the combination airfoil is zero. Hence the drag of the reversed airfoil must be equal to the drag of the airfoil in its initial direction.

Since, furthermore, the angle of attack of the plane infinitely thin airfoil does not change by the reversal, and since this angle is equal to the drag-lift ratio, it follows that for such special airfoils having a drag the total lift in both cases must also be the same. Hence, the lift slope must remain unchanged by the reversal for any airfoil.

Two airfoils may, of course, be associated with equal drag and with equal lift slope without being the reversal of each other. If the airfoils are the reversal of each other, no further proof is needed for equal drag and lift slope. The airfoils are not truly reversals of each other if the leading edge of the profile remains the leading edge unless the profiles are symmetric fore and aft. A full bodily reversal of the airfoil is necessary to make the preceding proof applicable. It is, on the other hand, apparent that the proof does not call for a single airfoil. Any assembly or combination of several airfoils in multiplane or in tandem relations or both, is permitted without invalidating the preceding proof. The airfoils must then not be reversed individually, but in the whole formation.

The proof does not lead to any information about the center of pressure of the reversed airfoil to its lift, unless the airfoil has zero camber. The symmetric relation holds only for the lift slope of an airfoil with arbitrary camber.

The argument of paragraph 10 implies the requirement that both the leading edge and the trailing edge be defined and serve as the separation line and as the rejoining line of the flow, respectively. Otherwise, very large streamwise increment velocity components would be generated, in contradiction with the assumptions made leading to the adoption of the linearized flow. The streamwise increment velocity component would not then be small throughout when compared with the velocity of advance. Hence the cancellation of the liftwise increment-velocity components of (a) and (b) in paragraph 9 would not be assured. Also the basic laws of mechanics would not be approximated by a solution obtained from the linearized differential equation. It is also implicit that the existence of a solution is not inconsistent with one double-edge condition. The requirement regarding the leading edge and the trailing edge is inconsistent with the existence of a solution of Laplace's equation in the theory of airfoils moving through an incompressible fluid. The requirement may be inconsistent with the existence of a solution of Eq. (1). It is by no means always inconsistent. It is not if the leading and the trailing edge are both "supersonic" throughout, that is, if the edges form an angle with the direction of advance larger than the

Mach angle. It is also not inconsistent if the local mean angle of attack of the airfoil is zero at all points. Supersonic edges throughout are not a necessary condition for the existence of a solution. There is some indication that nonexistence of a solution may require that portions of the leading edge be downstream and in the Mach direction relative to portions of the trailing edge. It seems that it has not yet been determined rigorously or finally under what conditions the two edges may be prescribed to serve as separate line and rejoining line, respectively, without excluding a solution of Eq. (1), and under what conditions they may not.

BIBLIOGRAPHY

L. Prandtl, "Theorie des Flugzeugtragflügels im zusammendrückbaren Medium," Luftfahrt-Forschung **13**, 313–319 (1936).

P. A. Lagerstrom, "Formulas in Three Dimensional Wing Theory," Douglas Aircraft Company, Report No. SM 11901 (July 8, 1946).

Th. von Karman, "Supersonic Aerodynamics," J. Aero. Sci. **14**, 373–402 (1947). See Section 5d.

The Impact of a Body on a Water Surface at an Arbitrary Angle

LEON TRILLING*

Hydrodynamics Laboratory, California Insitute of Technology, Pasadena, California

(Received August 18, 1949)

This paper presents an approximate method of determining the pressure distribution during impact on the surface of a body which strikes a horizontal water surface at an arbitrary angle.

The effect of the splash is neglected and the pressure on the free boundary is assumed proportional to the potential, as if the process were an impulse. The shape of the submerged portion of the striking body is approximated by a semi-ellipse (two dimensions), by a hemisphere, half an ellipsoid of revolution and half a general ellipsoid. Under those conditions, explicit results for the pressure distribution are found. As a special example, the impact pressures on a sphere striking at 45° angle are computed in detail.

AN estimate of the pressure forces on a body which strikes a water surface at an arbitrary angle is important in many problems, such as the design of seaplane hulls or naval ballistics.

The present work discusses the earliest stage of the water entry phenomenon which precedes the appearance of a cavity and displays the highest pressures. It is usually called the impact stage.

The problem of water impact has been discussed by numerous writers interested in the landing characteristics of seaplanes and the underwater trajectory of projectiles. Because of the difficulty of the problem, they were forced to various simplifying assumptions. Thus, von Karman[1] approximated the shape of the striking body by a growing flat plate in the plane of the water surface. H. Wagner[2,3] refined Karman's analysis by making some allowance for the water splash, and M. S. Plesset[4] by replacing the flat plate by an elliptic disk. To take into account the penetration of the body into the water, Schiffman and Spencer[5,6] approximated its shape by a spherical lens; P. Y. Chou[7] approximated it by a spherical bowl, whose potential was calculated by A. B. Basset.[8] Those latter investigations give accurate results in the case of normal impact, but they are too complicated to be generalized to impact at an arbitrary angle. Some simple two-dimensional problems were solved in all generality by L. I. Sedoff[9] who used the powerful methods of conformal transformation.

In the following pages, the problem is linearized and the classical methods of potential theory are applied to determine the pressure on two-dimensional bodies whose submerged portion may be approximated by a semi-ellipse, and on three-dimensional bodies which may be approximated by a hemisphere, half an ellipsoid of revolution, and half a general ellipsoid. Because of the linear nature of the simplified problem the effect of the vertical and horizontal components of impact velocity may be discussed separately and superimposed, so that one obtains results valid for any angle of impact.

1. STATEMENT OF THE PROBLEM

A three-dimensional half-space filled with incompressible fluid of density ρ at rest with a horizontal free

* The major portion of this study was carried out while the writer was with the Naval Ordnance Test Station; the writer also wishes to acknowledge the support of the ONR.

[1] Th. von Karman, "The impact on sea-plane floats during landing," NACA-TN 321 (1929).

[2] H. Wagner, "On the landing of seaplanes," ZFM (January, 1931)—NACA-TM 622.

[3] H. Wagner, "Uber stoss und gleitvorgange an der oberflache von flussigkeiten," ZAMM, **12**, 193–215 (1932).

[4] M. S. Plesset, "An investigation of the impact forces on torpedoes entering water," Douglas Company Report SM3937 (December, 1942).

[5] Schiffman and Spencer, "The force of impact on a sphere striking a water surface," AMP-421R-AMG-NYU No. 105 (February, 1945).

[6] Schiffman and Spencer, "The force of impact on a sphere striking a water surface (2)," AMP-42.2R-AMG-NYU No. 133 (July, 1945).

[7] P. Y. Chou, "On impact of spheres upon water," U.S. NOTS (1946), (to be published).

[8] A. B. Basset, "On the potential of an electrified spherical bowl and on the velocity potential due to the motion of an infinite liquid about such a bowl," Proc. Math. Soc. London, 1st Series **16**, 286–306 (1885).

[9] L. I. Sedoff, "The impact of a solid body which moves at the surface of an incompressible fluid," TCAHI Report No. 187, Moscow (1934).

surface, is struck by a body S which travels at a velocity $\bar{V}$, at an angle α from the horizontal plane. The fluid in the vicinity of the impact is set into motion; a small amount rises above the horizontal plane and forms a splash; a larger quantity is disturbed below the free surface, which is no longer horizontal, to make room for the body. The force of gravity, the hydrostatic force, viscous drag forces, surface tension forces, and the hydrodynamic forces all act on the body. But if the velocity $\bar{V}$ is large, during the impact stage, the hydrodynamic forces alone are significant, and the other forces are neglected in all that follows.

The hydrodynamic force is equal to the rate of change of momentum of the fluid set into motion by the impact. Since the fluid is assumed to be non-viscous and the motion is started from rest, the flow is irrotational and the velocity field may be derived from a potential Φ. The force is therefore:

$$\bar{F}=\frac{\rho}{\bar{V}}\frac{d}{dt}\int_{S+S'+S''}\Phi\frac{\partial\Phi}{\partial n}dS. \tag{1.1}$$

The integration is carried out over the boundaries of the fluid.** The forces acting on the body are therefore computed by a quadrature if the flow potential is known. Since the impact occurs in a very short time interval, $\bar{V}$ is assumed constant.

The velocity potential Φ of an incompressible perfect fluid satisfies the Laplace equation:

$$\Delta\Phi=0 \tag{1.2}$$

and certain conditions along the boundary of the fluid. The disturbance due to the impact must disappear far from the point of impact. Thus, if r denotes the distance from the point of impact, one has

$$\lim_{r\to\infty}\nabla\Phi=\lim_{r\to\infty}(\partial\Phi/\partial t)=0 \tag{1.3}$$

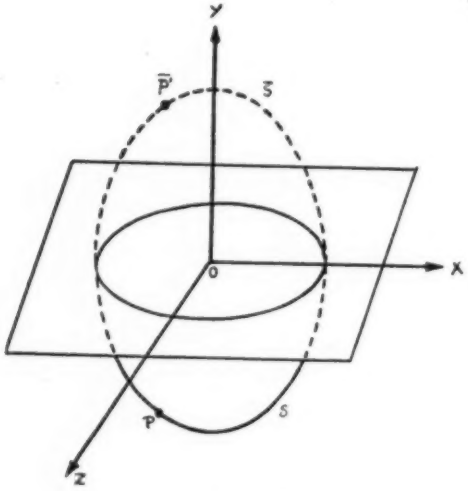


FIG. 1. General boundary conditions.

** S represents the surface of the submerged portion of the body, S' the free water surface, and S'' is a hemisphere of infinitely large radius with center at the body.

or the equivalent condition

$$\lim_{r\to\infty}\Phi=0. \tag{1.3a}$$

The velocity component normal to the surface of the impinging body must vanish:

$$\frac{\partial\Phi}{\partial n}\Big)_S=-V\cos\lambda, \tag{1.4}$$

where n is the direction normal to the wetted surface S, and λ is the angle between that normal and the direction of $\bar{V}$. Finally, along the free surface, the pressure in the fluid must equal atmospheric pressure, which may be taken equal to zero. This gives a condition on Φ, when Euler's equation is used:

$$(\partial\bar{u}/\partial t)+\bar{u}\cdot\nabla\bar{u}=(-1/\rho)\nabla p, \tag{1.5}$$

where $\bar{u}$ is the velocity vector in the fluid and p the pressure. An integration brings this equation into the form

$$-\partial\Phi/\partial t+\tfrac{1}{2}(\nabla\Phi)^2=-(p/\rho)+F(t). \tag{1.5a}$$

Boundary condition (1.3) applied on the free surface where p vanishes, gives $F(t)\equiv 0$, so that the boundary condition valid on the free surface is:

$$\tfrac{1}{2}(\nabla\Phi)^2-(\partial\Phi/\partial t)=0. \tag{1.5b}$$

One is thus led to investigate solutions of (1.2) subject to boundary conditions (1.3a, 1.4, 1.5b). Since (1.5b) is not linear, and the surface where it holds must be determined from the solution, this problem is extremely difficult; indeed, it has not been solved.

In linearizing boundary condition (1.5b), one argues that the impact occurs during a very short time interval, and may therefore be considered as an impulse. But it is known[10] that in the case of impulsive motion, the dynamic pressure is proportional to the velocity potential. Since the pressure vanishes on the free surface, one replaces condition (1.5b) by the more restrictive condition

$$\Phi(S')=0, \tag{1.5c}$$

where S' denotes the free surface. It is further assumed that the effect of the deformation of the free surface due to the splash is small, so that (1.5c) is applied along the original free surface:

$$\Phi(x,0,z,t)=0 \tag{1.5d}$$

in a cartesian system with origin at the center of impact, where the y axis is normal to the undisturbed free surface and the vector $\bar{V}$ lies in the plane $z=0$.

The problem of finding the potential Φ which satisfies Eq. (1.2) with boundary conditions (1.3a, 1.4, 1.5d) is a linear harmonic problem. It is convenient to investigate the potentials due to the various components

[10] H. Lamb, *Hydrodynamics* (Dover Publications, New York, 1944), sixth edition, London, (1932).

of $\bar{V}$ separately and to superimpose the results. Thus:

$$\Phi = V_x\varphi_1 + V_y\varphi_2 + V_z\varphi_3 + \Omega_{yz}\varphi_4 + \Omega_{zx}\varphi_5 + \Omega_{xy}\varphi_6 \quad (1.6)$$

where $V_{x,y,z}$ represent the x, y, z components of the velocity vector $\bar{V}$ and Ω_{ij} represent rotation velocities about the axes. In the problems to follow, V_z, Ω_{ij} vanish identically, so that

$$\Phi = V(\varphi_1 \cos\alpha + \varphi_2 \sin\alpha). \quad (1.6a)$$

If the symmetric image $\bar{S}$ of the wetted surface S with respect to the plane $y=0$ is constructed, (Fig. 1) and the field of flow is continued into the upper half-space, it follows from condition (1.5d) and the harmonic character of Φ that:

$$\Phi(x, y, z, t) = -\Phi(x, -y, z, t). \quad (1.5e)$$

The velocity components u, v, w, parallel to the x, y, z, axes, therefore have the following symmetry properties:

$$u(x, y, z) = -u(x, -y, z) \quad (1.7a)$$

$$v(x, y, z) = v(x, -y, z) \quad (1.7b)$$

$$w(x, y, z) = -w(x, -y, z) \quad (1.7c)$$

or, if the velocities normal to the surface $S+\bar{S}$ at two symmetric points $P, \bar{P}$ are compared,

$$(\partial\varphi/\partial n)_P = -(\partial\varphi/\partial n)_{\bar{P}}. \quad (1.7d)$$

If λ_1, λ_2 denote the angle between the normal to the surface $S+\bar{S}$ and the x, y axes, it follows from (1.7d) that boundary conditions (1.5, 1.4) become, for the potentials φ_1, φ_2

$$-\partial\varphi_1/\partial n = \cos\lambda_1 (|y|/y) \quad (1.8a)$$

$$-\partial\varphi_2/\partial n = \cos\lambda_2 \quad (1.8b)$$

these conditions being valid in the entire space.

The problem is thus split into two parts. The velocity component normal to the horizontal free surface induces the same potential as that of a fully submerged body in translation. The velocity component parallel to the horizontal free surface induces the potential which would obtain if the fluid above the plane of symmetry were flowing with a velocity equal and opposite that of the fluid below, causing a vortex sheet along the plane of symmetry.

One is now ready to express the momentum integral in terms of φ_1 and φ_2:

$$\frac{\rho}{V}\int_S \Phi \frac{\partial\Phi}{\partial n} dS' = \rho V \Big[\mathbf{i} \cos^2\alpha \int_S \varphi_1 \frac{\partial\varphi_1}{\partial n} dS + \sin\alpha\cos\alpha \Big(\mathbf{i}\int_S \varphi_2 \frac{\partial\varphi_1}{\partial n} dS + \mathbf{j}\int_S \varphi_1 \frac{\partial\varphi_2}{\partial n} dS \Big) + \mathbf{j}\sin^2\alpha \int_S \varphi_2 \frac{\partial\varphi_2}{\partial n} dS \Big] = \mathbf{i}M_x + (\mathbf{i}+\mathbf{j})M_{xy} + \mathbf{j}M_y, \quad (1.9)$$

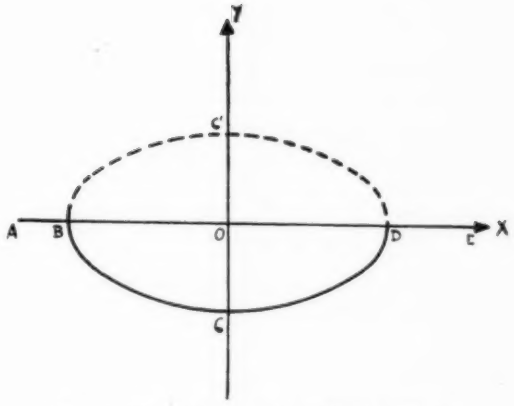


FIG. 2. Water impact of an elliptic cylinder.

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors parallel to the x, y axes. Note that φ_2 is positive over the entire surface S. On the other hand, φ_1 may be negative over part of the rear-half of the surface. If S is symmetric with respect to the plane $x=0$, φ_1 is antisymmetric and negative over the entire rear-half. But the dynamic pressure $p=\rho\Phi$ cannot become negative, since this would imply that the fluid is in tension and therefore separates from the body surface to form a cavity. Actually, because of the action of viscous forces and surface tension forces, separation is not instantaneous, and during the short period of time in which the impact occurs, the streamline pattern is not changed appreciably. But the integration in (1.9) must be carried out only over regions where $\varphi = \varphi_1\cos\alpha + \varphi_2\sin\alpha$ is positive.

The simplified problem formulated above does not depend on the time parameter directly since the equation of motion (1.2) and its boundary condition (1.3, 1.8) do not involve time variations. But as time passes, an increasing volume of the striking body is immersed in the fluid. Passage of time is, therefore, marked by a change in the dimensions and shape of the immersed body. Translation of the body in a horizontal direction does not change the shape of the immersed body and induces no force, but translation perpendicular to the free surface does.

In order to compute the time rate of change of momentum, therefore, it is necessary to determine how the immersed portion of the body changes with time. Most bodies whose impact may be of interest (projectiles, seaplane floats, etc.) have immersed portions of shapes so complicated that an analytic determination of φ_1, φ_2 is not possible. Following the scheme of previous calculations, one is led to approximate the immersed portion of the striking body at any time by some geometrically simple body (hemisphere, half-ellipsoid, etc.). The simple shape is defined by a certain number of parameters: the sphere by its radius, the ellipsoid of revolution by two axes, the general ellipsoid by three axes, etc. The approximating process consists in selecting certain properties of the actual body and defining the parameters of the approximating body so

that those properties are preserved. One can generally preserve as many properties as there are independent parameters defining the approximating body. For instance, if at any time t, the immersed portion of the striking body is approximated by a half-ellipsoid, the three axes of the ellipsoid are chosen so that the depth of penetration, the length and the immersed volume of the half-ellipsoid equal those of the actual body. As time passes, the depth of penetration, length and immersed volume of the actual body change, and the three axes of the approximating half-ellipsoid change correspondingly. The three axes of the ellipsoid are, therefore, known functions of time, when the motion of the actual body is known, either from measurements or from calculations. The momentum integral M (1.9a) of the approximating body depends only on its geometrical parameters $\alpha_1, \alpha_2, \alpha_3 \cdots \alpha_i$, which are known functions of time, since V and ρ are constant. The drag force is therefore given by:

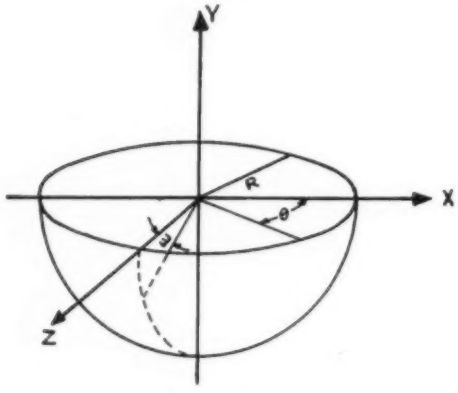


FIG. 3. Water impact of a sphere.

$$F_x=\rho V\cos\alpha\sum(d\alpha_i/dt)\times[(\partial M_x/\partial\alpha_i)\cos\alpha+(\partial M_{xy}/\partial\alpha_i)\sin\alpha] \quad (1.10a)$$

$$F_y=\rho V\sin\alpha\sum(d\alpha_i/dt)\times[(\partial M_{xy}/\partial\alpha_i)\cos\alpha+(\partial M_y/\partial\alpha_i)\sin\alpha]. \quad (1.10b)$$

The problem of finding approximately the forces acting on a given body when it strikes a horizontal surface is, therefore, investigated as follows:

An approximating body is selected and its parameters α_i are computed as functions of time from the known behavior of the actual body.

The potential problem for the approximating body is solved and the momentum integral is computed in terms of α_i.

The force function is found by application of Eqs. (1.10.)

The second step of this procedure is the most important and difficult. It can be carried out once and for all for several approximating body shapes. Then the results can be applied to a wide variety of problems by the proper fitting procedure.

In the following, the potential problem is solved for an infinitely wide semi-elliptic cylinder (two-dimensional problem solved by Sedoff's methods) and for a sphere, an ellipsoid of revolution and a general ellipsoid (solved by harmonic analysis methods). As an example of the fitting process, the water entry forces on a sphere striking at 45° are computed by using a general ellipsoid as the approximating body.

2. INFINITE SEMI-ELLIPTIC CYLINDER

The first potential problem discussed in detail concerns an infinitely wide semi-elliptic cylinder of major axis $2a$ and minor axis $2b$, the free surface $y=0$ coinciding with the major axis (Fig. 2). The boundary conditions to be satisfied by the potentials φ_1, φ_2 in the entire x, y plane are obtained from (1.8). If a stream function ψ is introduced, then, on the body surface,

$$\partial\varphi/\partial n=\partial\psi/\partial s=\cos\alpha(\partial y/\partial s)(|y|/y)-\sin\alpha(\partial x/\partial s), \quad (2.1)$$

where ds is a length element along the surface of the ellipse. This is integrated to give:

$$\psi=|y|\cos\alpha-x\sin\alpha. \quad (2.1a)$$

One is now required to find an analytic complex potential function $F(z)=F(x+iy)$ whose imaginary part satisfies (2.1a) on the boundary of the ellipse.

The conformal transformation

$$-2z=(a-b)\zeta+(a+b)1/\zeta \quad (2.2)$$

transforms the ellipse in the z plane into the unit circle in the ζ-plane. Therefore, in the ζ-plane the complex potential $F(\zeta)$ must satisfy the conditions below on the unit circle $\zeta=e^{i\theta}$:

$$\psi=b|\sin\theta|\cos\alpha+a\cos\theta\sin\alpha=\frac{2b\cos\alpha}{\pi}\left[1-2\sum_1^\infty\frac{\cos 2n\theta}{4n^2-1}\right]+a\sin\alpha\cos\theta. \quad (2.3)$$

It follows that:

$$F(\zeta)=\frac{2bi\cos\alpha}{\pi}\left[1+2\sum_1^\infty\frac{\zeta^{2n}}{4n^2-1}\right]+ai\sin\alpha\zeta. \quad (2.4)$$

To sum the power series inside and on the unit circle, note that

$$2\sum_1^\infty\frac{\zeta^{2n}}{4n^2-1}=\zeta\sum_1^\infty\frac{\zeta^{2n-1}}{2n-1}-\frac{1}{\zeta}\sum_1^\infty\frac{\zeta^{2n+1}}{2n+1} \quad (2.5a)$$

and since

$$\sum_1^\infty\zeta^{2n}=\zeta^2/1-\zeta^2 \quad (2.5b)$$

one has

$$\sum_1^\infty\frac{\zeta^{2n-1}}{2n-1}=\int_0^\zeta\frac{d\zeta}{1-\zeta^2}=\tfrac{1}{2}\log\frac{1+\zeta}{1-\zeta} \quad (2.5c)$$

$$\sum_1^\infty\frac{\zeta^{2n+1}}{2n+1}=\int_0^\zeta\frac{\zeta^2 d\zeta}{1-\zeta^2}=-\zeta+\tfrac{1}{2}\log\frac{1+\zeta}{1-\zeta} \quad (2.5d)$$

so that

$$F(\zeta)=\frac{ib\cos\alpha}{\pi}\left(\frac{1}{\zeta}-\zeta\right)\log\frac{1+\zeta}{1-\zeta}+ia\zeta\sin\alpha. \quad (2.6)$$

All the manipulations carried out above are valid in the annulus $0<|\zeta|<1$ because all series and integrals converge uniformly there. It is easily verified that $F(\zeta)$ is an analytic function regular everywhere in and on the unit circle. Separating real and imaginary parts, on the circle, one finds

$$\psi=b\cos\alpha|\sin\theta|+a\cos\theta\sin\alpha \quad (2.7a)$$

$$\varphi=-\sin\theta[(2b/\pi)\cos\alpha\log|\tan(\theta/2)|+a\sin\alpha]. \quad (2.7b)$$

The boundary condition (2.3) is therefore satisfied and the pressure distribution is given by (2.7b). The separation point occurs at θ_s defined by

$$(2b/\pi)\cos\alpha\log|\tan(\theta_s/2)|+a\sin\alpha=0 \quad (2.8a)$$

or

$$\theta_s=2\tan^{-1}\exp[-(a\pi/2b)\tan\alpha]. \quad (2.8b)$$

In most cases $a\gg b$ and $\tan\alpha=0(1)$; then θ_s is small: $\theta_s=0(10^{-3})$ and the separation phenomenon may be neglected. Under those circumstances, the momentum integrals are:

$$M_x=2b^2/\pi \quad (2.9a)$$

$$M_{xy}=M_{yx}=0 \quad (2.9b)$$

$$M_y=\pi a^2/2. \quad (2.9c)$$

3. THE SPHERE

Consider a sphere of radius R half-submerged in a fluid (Fig. 3). The potential $\varphi=\varphi_1\cos\alpha+\varphi_2\sin\alpha$ must satisfy the equation

$$\Delta\varphi_1=\Delta\varphi_2=0 \quad (3.1)$$

with the boundary conditions:

$$\lim_{r\to\infty}\varphi_1=\lim_{r\to\infty}\varphi_2=0 \quad (3.2)$$

$$(\partial\varphi_2/\partial r)(R,\theta,\omega)=-\cos\omega \quad (3.3a)$$

$$(\partial\varphi_1/\partial r)(R,\theta,\omega)=-(|\cos\omega|/\cos\omega)\cos\theta. \quad (3.3b)$$

The potential φ_2 is immediately found:

$$\varphi_2=(R^3/2r^2)\cos\omega. \quad (3.4)$$

The potential φ_1 is determined by fitting the general harmonic function convergent for large r

$$\varphi_2=\sum_0^\infty\sum_0^{n+1}\frac{1}{r^{n+1}}P_n^s(\cos\omega)[a_n^s\cos s\theta+b_n^s\sin s\theta] \quad (3.5)$$

to satisfy boundary condition (3.3b). It is clear by inspection that $b_n^s=0$ and $a_n^{s>1}=0$, while the non-vanishing coefficients a_n^1 must satisfy the relation:

$$\sum_0^\infty\frac{n+1}{R^{n+2}}a_n^1P_n^1(\cos\omega)=\sum_0^\infty\alpha_nP_n^1$$

$$=-1 \quad [0<|\omega|<(\pi/2)]$$

$$=1 \quad [(\pi/2)<|\omega|<\pi]. \quad (3.6)$$

If Eq. (3.6) is multiplied by $P_k^1(\cos\omega)\sin\omega$ and integrated from 0 to π, all terms in the series except the k

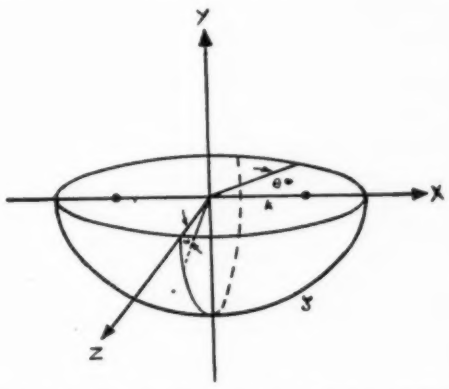


FIG. 4. Water impact of an ellipsoid of revolution.

term vanish because of the orthogonality of Legendre polynomials. One, therefore, obtains:

$$\alpha_k=\frac{2k+1}{2k(k+1)}\left[\int_0^{\pi/2}P_k^1(\cos\omega)\sin\omega d\omega-\int_{A/2}^{\pi}P_k^1(\cos\omega)\sin\omega d\omega\right]. \quad (3.7a)$$

Since $P_k^1(\cos\omega)=\sin\omega[dP_k(\cos\omega)/d(\cos\omega)]$, this is rewritten, with due regard to symmetry properties:

$$\alpha_{2k}=\frac{4k+1}{2k(2k+1)}\int_0^{\pi/2}\sin^2\omega P_{2k}'(\cos\omega)d\omega. \quad (3.7b)$$

$$\alpha_{2k+1}=0$$

By use of Rodrigues's Formula[11] the Legendre polynomial $P_{2k}(z)$ can be written as:

$$P_{2k}(z)=\frac{1}{2^kk!}\sum_{r=0}^{k}\binom{k-r}{r}\frac{(2k+2r-1)!!}{(2r-1)!!}(-)^{k-r}z^{2r}. \quad (3.8a)$$

It is also known[12] that

$$\int_0^{\pi/2}\sin^2\omega\cos^{2r-1}\omega d\omega=\frac{(2r-2)!!}{(2r+1)!!}. \quad (3.8b)$$

[11] Whittaker and Watson, *Modern Analysis* (Cambridge University Press, London, 1940), fourth edition.

[12] H. B. Dwight, *Tables of Integrals and Other Mathematical Data* (MacMillan Company, Ltd., London, 1934).

Combining formulas (3.8a) and (3.8b), one obtains:

$$\alpha_{2k}=\frac{4k+1}{2^k k(k!)(2k+1)}\sum_{r=1}^{k}(-)^{k-r+1}\times\binom{k-r}{k}\frac{r(2r-2)!!(2k+2r-1)!!}{(2r-1)!!(2r+1)!!}. \quad (3.9)$$

Thus, for instance:

$$\begin{aligned}&\alpha_2=5/6=0.833 && \alpha_8=-0.041\\ &\alpha_4=-(3/40)=-0.075 && \alpha_{10}=0.0658.\\ &\alpha_6=0.147 && \alpha_{12}=-0.0094\cdots\end{aligned} \quad (3.9a)$$

The convergence of the series (3.6) is established by noting that the integral in (3.7b) can be determined asymptotically as $k\to\infty$ as follows:

$$\begin{aligned}\lim_{k\to\infty}\int_0^{\pi/2}P_{2k}'(\cos\omega)\sin^2\omega d\omega&=\int_0^{\pi/2}\sin^2\omega\lim_{k\to\infty}P_{2k}'(\cos\omega)d\omega\\&=\int_0^{\pi/2}\sin^2\omega(k/\pi)^{\frac{1}{2}}\sin 2k\omega d\omega\\&=\tfrac{1}{4}(k/\pi)^{\frac{1}{2}}\int_0^{\pi}(1-\cos\theta)\sin k\theta d\theta\sim 1/(k)^{\frac{1}{2}}.\end{aligned} \quad (3.10)$$

The general coefficient α_k therefore behaves as $k^{-\frac{1}{2}}$. Since[13]

$$\lim_{k\to\infty}P_{2k}^{1}(\cos\theta)=-2(k/\pi\sin\theta)^{\frac{1}{2}}\sin[(2k+\tfrac{1}{2})\theta+3\pi/4] \quad (3.10')$$

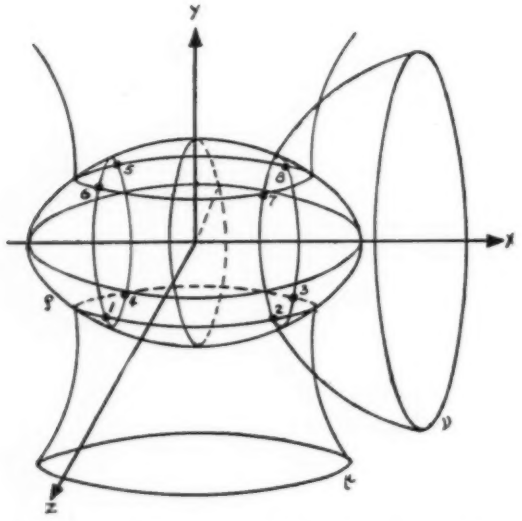


Fig. 5. Coordinate system for a general ellipsoid.

[13] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, London, 1931), Chapter 11.

the series (3.6) behaves as $\sum(e^{ik\theta}/k)$ which is known to converge.

The potential of the horizontal flow past a sphere of radius R is, therefore,

$$\varphi_1=R\cos\theta\sum_1^{\infty}(R/r)^{n+1}(\alpha_n/n+1)P_n^1(\cos\omega). \quad (3.11)$$

The separation point is reached along the curve $\theta(\omega)$ where the following relation holds:

$$\theta=\sec^{-1}[-2\cot\alpha\sec\omega\sum_1^{\infty}(\alpha_n/n+1)P_n^1(\cos\omega)]. \quad (3.12)$$

If separation effects are neglected, one finds the following momentum integrals:

$$M_x=(\pi/2)\rho R^3\sum_0^{\infty}(\alpha_n)^2 \quad (3.13a)$$

$$M_{xy}=M_{yz}=0 \quad (3.13b)$$

$$M_y=(\pi/3)\rho R^3. \quad (3.13c)$$

4. HALF AN ELLIPSOID OF REVOLUTION

A slight generalization is obtained if the hemisphere considered above is replaced by an ellipsoid of revolution whose axis is parallel to the horizontal velocity component (Fig. 4). The potential problem is most simply discussed in terms of elliptic coordinates. If x, y, z, are cartesian coordinates as before, and the distance between the foci of the ellipsoid is $2k$, the new coordinates μ, ζ, ω are defined by

$$x=k\mu\zeta \quad (4.1a)$$

$$y=k[(1-\mu^2)(\zeta^2-1)]^{\frac{1}{2}}\sin\omega \quad (4.1b)$$

$$z=k[(1-\mu^2)(\zeta^2-1)]^{\frac{1}{2}}\cos\omega. \quad (4.1c)$$

The ellipsoid under study is defined by a value ζ_0 of the parameter ζ, and solutions are sought in the domain $\zeta>\zeta_0$ which corresponds to the space outside the ellipsoid.

The boundary conditions corresponding to (1.8a, b) are here:

$$\left(\frac{\partial\varphi_1}{\partial\zeta}\right)_{\zeta_0}=-k\mu\frac{|\cos\omega|}{\cos\omega} \quad (4.2a)$$

$$\left(\frac{\partial\varphi_2}{\partial\zeta}\right)_{\zeta_0}=-k\zeta_0\left(\frac{1-\mu^2}{\zeta_0^2-1}\right)^{\frac{1}{2}}\sin\omega. \quad (4.2b)$$

The potential φ_2 is known to be[11]

$$\varphi_2=\frac{k}{\frac{1}{2}\log(\zeta_0+1/\zeta_0-1)-[\zeta_0^2-2/\zeta_0(\zeta_0^2-1)]}\times\left\{\tfrac{1}{2}\log\frac{\zeta+1}{\zeta-1}-\frac{\zeta}{\zeta^2-1}\right\}[(1-\mu^2)(\zeta^2-1)]^{\frac{1}{2}}\cos\omega. \quad (4.3)$$

The general potential convergent for $\zeta > \zeta_0$ is

$$\varphi_1 = \sum_0^\infty \sum_1^{n+1} Q_n^s(\zeta) P_n^s(\mu) [a_n^s \sin s\omega + b_n^s \cos s\omega] \quad (4.4)$$

so that boundary condition (4.2a) gives:

$$\sum_0^\infty \sum_1^{n+1} Q_n'^s(\zeta_0) P_n^s(\mu) a_n^s \sin s\omega$$

$$= \sum_0^\infty \sum_1^{n+1} \alpha_n^s P_n^s(\mu) \sin s\omega = -\mu (-\pi < \omega < 0) \quad (4.5a)$$

$$= \mu (0 < \omega < \pi)$$

from which the coefficients α_n^s are determined. Multiplication of (4.5a) by $\sin(r\omega)$ and integration between $-\pi$ and π leads to:

$$\sum_0^\infty \alpha_n^{2r-1} P_n^{2r-1}(\mu) = -\frac{4}{\pi} \frac{(-)^r \mu}{2r+1} \quad (4.5b)$$

$$\alpha_n^{2r} = 0.$$

Multiplication of (4.5b) by $P_k^{2r+1}(\mu)$ and integration between -1 and 1 leads to:

$$\alpha_{2k}^{2r-1} = \frac{(-)^r 8(4k+1)(2k-2r-1)!}{\pi(2r+1)(2k+2r+1)!} \times \int_0^1 \mu P_{2k}^{2r+1}(\mu) d\mu \quad (4.5c)$$

$$\alpha_{2k+1}^{2r+1} = 0.$$

After two integrations by parts, the integral in (4.5c) becomes:

$$\int_0^1 \mu P_{2k}^{2r+1}(\mu) d\mu = \int_0^1 \mu (1-\mu^2)^{\frac{1}{2}} \frac{d}{d\mu} P_{2k}^{2r}(\mu) d\mu$$

$$= \int_0^1 \frac{2\mu^2 - 1}{(1-\mu^2)^{\frac{1}{2}}} P_{2k}^{2r}(\mu) d\mu$$

$$= \int_0^1 (2k^2 - 1) \frac{d}{d\mu} P_{2k}^{2r-1}(\mu) d\mu$$

$$= -4 \int_0^1 \mu P_{2k}^{2r-1}(\mu) d\mu \quad (4.6a)$$

since

$$P_{2k}^{2r+1}(0) = P_{2k}^{2r+1}(1) = 0.$$

The recursion formula (4.6a) gives at once:

$$\int_0^1 \mu P_{2k}^{2r+1}(\mu) d\mu = (-4)^r \int_0^1 \mu P_{2k}^1(\mu) d\mu$$

$$= (-4)^r \int_0^{\pi/2} \sin^2\theta P_{2k}'(\cos\theta) d\theta, \quad (4.6b)$$

By using the representation (3.8a) for $P_{2k}(z)$ and the relation:

$$\int_0^{\pi/2} \sin^2\theta \cos^{2p}\theta d\theta = \frac{\pi}{2} \frac{(2p-1)!!}{(2p+2)!!} \quad (4.6c)$$

one finds the coefficient α_{2k}^{2r+1}

$$\alpha_{2k}^{2r+1} = \frac{4^{r+1}}{2^{k-1}k!} \frac{4k+1}{2r+1} \frac{(2k-2r-1)!}{(2k+2r+1)!} \times \sum_{p=1}^{k} (-)^{k-p+1} \binom{k-p}{p} \frac{p(2k+2p-1)!!}{(2p+2)!!} \quad (4.7)$$

Thus for instance:

$$\begin{array}{lll} \alpha_2^1 = 1.250 & & \\ \alpha_4^1 = 0.281 & & \\ \alpha_6^1 = 0.127 & \alpha_4^3 = 0.00161 & \alpha_6^5 = 0.00001. \\ \alpha_8^1 = 0.0726 & \alpha_6^3 = 0.00002 & \\ \alpha_{10}^1 = 0.0471 & & \\ \alpha_{12}^1 = 0.0328 & & \end{array} \quad (4.7a)$$

The uniform convergence of the doubly infinite series (4.5a) is established by the Weierstrass M test if $\sum\sum \alpha_k^r$ converges: The integral in (4.5c) converges:

$$\lim_{k \to \infty} \left| \int_0^1 \mu P_{2k}^1 d\mu \right| < \frac{A}{(k)^{\frac{1}{2}}} \quad (4.8a)$$

as can be shown by an argument similar to that of (3.10). For any value of r, therefore, the series

$$\sum_{k=0}^\infty \alpha_k^r = S'(r)$$

converges absolutely. For any given value of k,

$$\lim_{r \to \infty} \sum_0^k \alpha_k^r < A(k) \sum \frac{4^r}{(2r+1)(4r)!} \quad (4.8b)$$

so that the sum of rows as well as the sum of columns converges absolutely. Then, one has

$$|\sum\sum \alpha_k^r| < \sum\sum (1/(k+r)!) \quad (4.8c)$$

which establishes the convergence of the series. The potential φ_1 for the horizontal flow past the half-ellipsoid of revolution is, therefore:

$$\varphi_1 = \sum_{n=1}^\infty \sum_{s=0}^{n-1} \frac{Q_{2n}^{2s+1}(\zeta)}{Q_{2n}'^{2s+1}(\zeta_0)} \times \alpha_{2n}^{2s+1} P_{2n}^{2s+1}(\mu) \sin(2s+1)\omega. \quad (4.9)$$

The momentum integrals are:

$$M_x = -\frac{\pi}{4} k^3 (\zeta_0^2 - 1) \sum_{n=1}^\infty \sum_{s=1}^{n-1} \frac{Q_{2n}^{2s+1}(\zeta_0)}{Q_{2n}'^{2s+1}(\zeta_0)} \times \frac{(2n+2s+1)!}{(2n-2s-1)!} \frac{(\alpha_{2n}^{2s+1})^2}{4n+1} \quad (4.10a)$$

$$M_{xy}=M_{yz}=0 \tag{4.10b}$$

$$M_y=\frac{\pi k^3}{2}\zeta_0(\zeta_0^2-1)$$

$$\times\frac{\log(\zeta_0+1/\zeta_0-1)-(2\zeta_0/\zeta_0^2-1)}{\log(\zeta_0+1/\zeta_0-1)-[2(\zeta_0^2-2)/\zeta_0(\zeta_0^2-1)]}. \tag{4.10c}$$

Note that the entire argument was carried through for a prolate ellipsoid ($\zeta>1$). If the ellipsoid is oblate ($\zeta<1$), ζ is replaced by $i\zeta$ and the function $q(\zeta)$ replaces $Q(\zeta)$, where

$$g_n^s(\zeta)=Q_n^s(i\zeta) \tag{4.11}$$

everything else remaining unchanged.

5. HALF A GENERAL ELLIPSOID

A further generalization is obtained by considering an ellipsoid with focal length $2k$ parallel to the horizontal velocity component, and focal length $2h$ horizontal and normal to it. The problem is now formulated in terms of ellipsoidal coordinates defined by E. W. Hobson:[13]

$$x=\rho\mu\nu/hk \tag{5.1a}$$

$$y=\frac{[(\rho^2-h^2)(\mu^2-h^2)(h^2-\nu^2)]^{\frac{1}{2}}}{h(k^2-h^2)^{\frac{1}{2}}} \tag{5.1b}$$

$$z=\frac{[(\rho^2-k^2)(k^2-\mu^2)(k^2-\nu^2)]^{\frac{1}{2}}}{k(k^2-h^2)^{\frac{1}{2}}}. \tag{5.1c}$$

The surfaces $\rho=$const. represent confocal ellipsoids; ρ_0 in particular represents the ellipsoid under investigation; $\mu=$const. are hyperboloids of one sheet and $\nu=$const. are hyperboloids of two sheets. The coordinate system is shown on Fig. 5. Any point is defined uniquely if the signs of the radicals in (5.1) are defined. Boundary conditions (1.8a, b) become in this system:

$$\frac{\partial\varphi_1}{\partial\rho}\Big)_{\rho_0}=-\frac{\mu\nu}{hk}\frac{|y|}{y} \tag{5.2a}$$

$$\frac{\partial\varphi_2}{\partial\rho}\Big)_{\rho_0}=-\frac{\rho_0}{h}\left\{\frac{[(\mu^2-h^2)(h^2-\nu^2)]}{[(\rho_0^2-h^2)(k^2-h^2)]}\right\}^{\frac{1}{2}}. \tag{5.2b}$$

The potential φ_2 is given explicitly in the standard literature[11] by the formula:

$$\varphi_2=\frac{\rho_0}{2h}\left[\frac{(\rho_0^2-h^2)(\rho_0^2-k^2)}{k^2-h^2}\right]^{\frac{1}{2}}$$

$$\times\frac{1}{1-\rho_0[(\rho_0^2-h^2)(\rho_0^2-k^2)]^{\frac{1}{2}}\displaystyle\int_{\rho_0}^{\infty}\frac{d\rho}{(\rho^2-h^2)^{\frac{1}{2}}(\rho^2-k^2)^{\frac{1}{2}}}}$$

$$\times[(\rho^2-h^2)(\mu^2-h^2)(h^2-\nu^2)]^{\frac{1}{2}}$$

$$\times\int_{\rho}^{\infty}\frac{d\rho}{(\rho^2-h^2)^{\frac{1}{2}}(\rho^2-k^2)^{\frac{1}{2}}}. \tag{5.3}$$

The general potential convergent for $\rho>\rho_0$ is

$$\varphi_1=\sum_0^\infty\sum_1^{2n+1}a_n^s F_n^s(\rho)E_n^s(\mu)E_n^s(\nu), \tag{5.4a}$$

where F_n^s, E_n^s are general Lamé functions, convergent as $\rho\to\infty$ and μ, $\nu\to 0$, and satisfying Lamé's differential equation:

$$(\lambda^2-h^2)(\lambda^2-k^2)(d^2E_n/d\lambda^2)$$
$$+\lambda(2\lambda^2-h^2-k^2)(dE_n/d\lambda)$$
$$+[(h^2+k^2)p-n(n+1)\lambda^2]E_n=0 \tag{5.4b}$$
$$\lambda=\rho,\ \mu,\ \nu$$

the parameter n is an integer, while $p(h, k, n)$ is chosen so that E_n^s is a polynomial. The theory of Lamé functions is discussed in detail in.[13]

Boundary condition (5.2a) gives here:

$$hk\sum_0^\infty\sum_1^{2n+1}F_n'^s(\rho_0)a_n^sE_n^s(\mu)E_n^s(\nu)$$

$$=\sum_0^\infty\sum_1^{2n+1}\alpha_n^sE_n^s(\mu)E_n^s(\nu)=\mu\nu[(\mu^2-h^2)^{\frac{1}{2}}>0]$$

$$=-\mu\nu[(\mu^2-h^2)^{\frac{1}{2}}<0]. \tag{5.5}$$

From the form of the function to be expanded, it is clear that only odd Lamé functions K_{2n+1}^s of the form

$$K_{2n+1}^s(\lambda)=\beta_1^s\lambda+\beta_3^s\lambda^3+\cdots\beta_{2n+1}^s\lambda^{2n+1} \tag{5.6a}$$

will appear in φ_1. The coefficients β_n^s are defined by recurrence formulas and depend on (h, k, p). They have been tabulated by Guerritore.[14] It is then known[15] that the coefficients α_n^s are given by:

$$\alpha_{2n+1}^s=\frac{\displaystyle\int_0^h\int_h^k(\mu^2-\nu^2)\mu\nu K_{2n+1}^s(\mu)K_{2n+1}^s(\nu)\frac{d\mu}{[(h^2-\mu^2)(\mu^2-k^2)]^{\frac{1}{2}}}\frac{d\nu}{[(h^2-\nu^2)(k^2-\nu^2)]^{\frac{1}{2}}}}{\displaystyle\int_0^h\int_h^k(\mu^2-\nu^2)[K_{2n+1}^s(\mu)K_{2n+1}^s(\nu)]^2\frac{d\mu}{[(h^2-\mu^2)(\mu^2-k^2)]^{\frac{1}{2}}}\frac{d\nu}{[(h^2-\nu^2)(k^2-\nu^2)]^{\frac{1}{2}}}}. \tag{5.7}$$

[14] M. Guerritore, Table of Lamé functions, Giornale de Matematica (2) XVI, 164–172 (1909).
[15] O. Volk, "Uber die entwicklung von functionen zweier complexen veranderlichen nach lameschen functionen," Math. Zeitsch 23, 224 (1925).

If the expansion (5.6) for $K_{2n+1}{}^{s}(\lambda)$ is substituted into the integrals (5.7), it is found that one must investigate integrals of the form:

$$A_{2r}(\lambda)=\int_0^{h}\frac{\lambda^{2r}d\lambda}{[(h^2-\lambda^2)(k^2-\lambda^2)]^{\frac{1}{2}}} \tag{5.8a}$$

$$B_{2r}(\lambda)=\int_h^{k}\frac{\lambda^{2r}d\lambda}{[(\lambda^2-h^2)(k^2-\lambda^2)]^{\frac{1}{2}}}. \tag{5.8b}$$

It is easily verified that:

$$A_0=(1/k)F(h/k) \qquad B_0=(1/k)F\left[\frac{(k^2-h^2)^{\frac{1}{2}}}{k}\right] \tag{5.9a,b}$$

$$A_2=k(F(h/k)-E(h/k)) \quad B_2=kE\left[\frac{(k^2-h^2)^{\frac{1}{2}}}{k}\right]. \tag{5.9c,d}$$

Repeated integrations by parts then lead to the recurrence formulas:

$$(2n-1)A_{2n}=(2n-4)(h^2+k^2)A_{2n-2} -(2n-3)h^2k^2A_{2n+4} \tag{5.10}$$

and an identical result for B_{2n}. The integrals A_{2n}, B_{2n} are, therefore, expressed as functions of the parameters h, k in terms of elliptic integrals.

The numerator and denominator of (5.7) are now simply expressed in terms of the known coefficients $\beta_n{}^{s}$ and the integrals A_{2n}, B_{2n}.

$$N_{2n+1}{}^{s}=(\beta_1{}^{s})^2(B_4A_2-B_2A_4) +\beta_1{}^{s}\beta_3{}^{s}(B_6A_2-A_6B_2) +(\beta_3{}^{s})^2(B_6A_4-B_4A_6)\cdots \tag{5.11a}$$

$$D_{2n+1}{}^{s}=(\beta_1{}^{s})^4(B_4A_2-B_2A_4) +2\beta_3{}^{s}(\beta_1{}^{s})^3(B_6A_2-A_6B_2)\cdots \tag{5.11b}$$

$$\alpha_{2n+1}{}^{s}=N_{2n+1}{}^{s}/D_{2n+1}{}^{s}. \tag{5.11c}$$

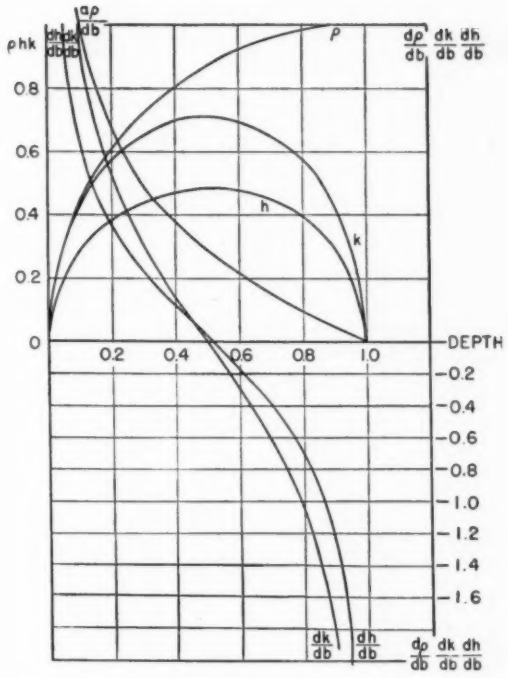


FIG. 6. The approximating functions for a partially submerged sphere and their derivatives.

The potential function φ_1 is then given by:

$$\varphi_1=\frac{1}{hk}\sum_0^{\infty}\sum_1^{n+1/2}\frac{F_{2n+1}{}^{s}(\rho)}{F_{2n+1}{}'^{s}(\rho_0)} \times\alpha_{2n+1}{}^{s}K_{2n+1}{}^{s}(\mu)K_{2n+1}{}^{s}(\nu). \tag{5.12}$$

The momentum integrals are:

$$M_x=\frac{[(\rho_0^2-h^2)(\rho_0^2-k^2)]^{\frac{1}{2}}}{h^2k^2}\sum_0^{\infty}\sum_1^{n+1/2}\frac{F_{2n+1}{}^{s}(\rho_0)}{F_{2n+1}{}'^{s}(\rho_0)}(\alpha_{2n+1}{}^{s})^2D_{2n+1}{}^{s} \tag{5.13a}$$

$$M_{xy}=M_{yx}=0 \tag{5.13b}$$

$$M_y=(\pi\rho_0/12)[(\rho_0^2-h^2)(\rho_0^2-k^2)]^{\frac{1}{2}}\frac{\rho_0[(\rho_0^2-h^2)(\rho_0^2-k^2)]^{\frac{1}{2}}\int_{\rho_0}^{\infty}\frac{d\rho}{\rho_0(\rho^2-h^2)^{\frac{1}{2}}(\rho^2-k^2)^{\frac{1}{2}}}}{1-\rho_0[(\rho_0^2-h^2)(\rho_0^2-k^2)]^{\frac{1}{2}}\int_{\rho_0}^{\infty}\frac{d\rho}{\rho_0(\rho^2-h^2)^{\frac{1}{2}}(\rho^2-k^2)^{\frac{1}{2}}}}. \tag{5.13c}$$

The first three terms of this expansion were computed numerically and used in Section 6.

6. WATER IMPACT OF A SPHERE AT AN ENTRY ANGLE OF 45°

The results of the previous pages are now applied to a specific problem. The drag on a sphere of radius R which strikes the water surface at 45° is computed as a function of time. The submerged portion of the sphere is approximated by a half-ellipsoid in such a manner that the depth of penetration d of the sphere equals the minor axis, the diameter of the circle of intersection between the sphere and the free surface equals the major axis and the submerged volume of the sphere, V, equal the volume of the semi-ellipsoid. If the dimensionless length ξ is introduced, one has:

$$\xi=tV\sin\alpha/R \tag{6.1a}$$

$$\rho_0(t)/R=[\xi(2-\xi)]^{\frac{1}{2}} \tag{6.1b}$$

$$k(t)/R=[2\xi(1-\xi)]^{\frac{1}{2}} \tag{6.1c}$$

$$h(t)/R=\tfrac{1}{2}\left[\frac{\xi(1-\xi)(7-3\xi)}{2-\xi}\right]^{\frac{1}{2}} \tag{6.1d}$$

$$d\alpha_i/dt=(V\sin\alpha/R)(d\alpha_i/d\xi). \tag{6.1e}$$

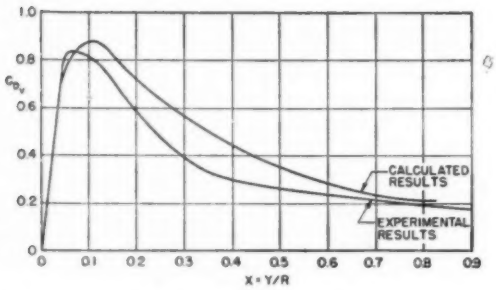


FIG. 7. Vertical component of drag of a sphere of unit radius in water impact $C_D = F_Y/\pi R^2\rho(V^2/2)$.

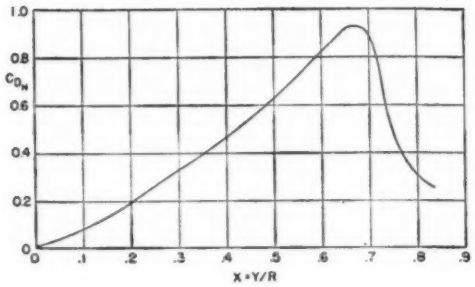


FIG. 8. Horizontal component of drag of a sphere in water impact.

The force functions are given by formulas (1.10a, b) and the derivatives are given explicitly as:

$$d\rho_0/dt = V\sin\alpha\frac{1-\xi}{2[\xi(2-\xi)]^{\frac{1}{2}}} \tag{6.2a}$$

$$dk/dt = V\sin\alpha\frac{1-2\xi}{[2\xi(1-\xi)]^{\frac{1}{2}}} \tag{6.2b}$$

$$dh/dt = V\sin\alpha\frac{7-20\xi+14\xi^2-3\xi^3}{2[\xi(1-\xi)(7-3\xi)(2-\xi)^3]^{\frac{1}{2}}} \tag{6.2c}$$

while the derivatives $\partial M/\partial\alpha_i$ are obtained by differentiating (5.13) with respect to ρ_0, h, k. This formulation is valid only while $\xi<1$, or up to the time when the sphere is half-submerged. After that instant, flow past an ellipsoid does not give a good approximation to the streamline pattern, and a cavity is beginning to develop.

The results of the calculation for the sphere are shown on Figs. 6, 7, 8. It is found that the value of the vertical drag component agrees fairly well with the value obtained in unpublished experiments carried out at the Morris Dam Hydrodynamics Station; it also gives the same type of result as that obtained by Schiffman and Spencer. The horizontal component of drag reaches its greatest value at a later point.

CONCLUSION

The impact of a body of arbitrary shape at an arbitrary angle was investigated by approximate methods. The two main approximations consisted in linearizing the boundary condition at the free surface by neglecting the splash and in replacing the submerged portion of the body at any time by an ellipsoid of identical depth of penetration, length, and submerged volume. As an illustration of the method, the impact of a sphere at a 45° angle was studied in some detail. The vertical force component turned out to be in good agreement with experimental results. The behavior of the horizontal component, which has not been measured, was also computed and shown.

Depolymerization by Ultrasonic Irradiation: The Role of Cavitation

ALFRED WEISSLER
Naval Research Laboratory, Washington, D. C.
(Received July 7, 1949)

Contrary to current belief, cavitation has been found responsible for the depolymerizing effect of intense ultrasonic waves. This was demonstrated by irradiating two portions of a 1 percent polystyrene solution in toluene under conditions identical except for the following. The first portion was given no special prior treatment, showed many cavitation bubbles during irradiation, and decreased in molecular weight (as measured by the intrinsic viscosity) to about one-tenth of the initial value. The second portion was given a preliminary treatment of degassing by boiling under vacuum, showed no cavitation bubbles during the irradiation, and underwent no appreciable change in molecular weight.

Similar experiments with solutions of hydroxyethyl cellulose in water showed that, in this case also, cavitation is necessary for depolymerization. The opposite conclusion of earlier investigators is attributed to their inadequate method for eliminating cavitation.

Oxidants known to be produced by ultrasonic waves in solutions containing dissolved oxygen or nitrogen cannot be responsible for the degradation, because substantially the same amount of depolymerization occurs even when helium is the only gas present.

INTRODUCTION

SOLUTIONS of high polymers may be affected by ultrasonic irradiation in two different ways.[1,2] First, Freundlich and Gillings found[3] that weak irradiation of fresh solutions of gelatin in water causes a considerable reduction in the structural or non-Newtonian viscosity. But on standing for several days, the solution regains almost as high a viscosity as it had before irradiation. This effect is believed due to the temporary breaking of the loose gel network of van der Waals bonds between adjacent polymer molecules. It does not take place to an appreciable extent in the absence of cavitation.[4]

The second way in which ultrasonic waves may affect high polymers is to cause depolymerization—the actual breaking of chemical bonds in the chain, to give molecules of smaller size than originally, and therefore a *permanent* decrease in viscosity. This was studied in detail by Schmid and Rommel[5] in order to determine whether cavitation is responsible for this effect also. In a typical experiment, they found that a solution of nitrocellulose in butyl acetate underwent considerable degradation when irradiated in a steel bomb for an hour under atmospheric pressure, but less degradation under 5 atmos. of nitrogen, and none at all under 8 atmos. of oxygen. Similarly, a solution of polystyrene in toluene suffered much depolymerization when irradiated under atmospheric pressure, less under 8 atmos. of O_2, and still less under 15 atmos. of O_2. Schmid concluded that depolymerization occurs in the complete absence of cavitation, because as little as 5 atmos. excess pressure would have suppressed cavitation caused by the moderate-intensity ultrasonic waves used.

However, this conclusion appears questionable, for the following reason. Although 5 atmos. excess hydraulic pressure applied in the body of the liquid by a piston will eliminate cavitation under the given conditions, applying 5 atmos. of oxygen above the surface will force 5 times as much oxygen into solution (Henry's Law). Then when the instantaneous pressure is reduced 1 or 2 atmos. in the negative part of the sound wave cycle, oxygen will come out of solution and bubbles will be formed in the liquid. Therefore, it seemed desirable to investigate this question further, using a more positive means for eliminating cavitation: degassing the solution by boiling under vacuum for at least 15 minutes, prior to irradiation.

EXPERIMENTAL

A sample of high molecular-weight polystyrene,[6] obtained through the courtesy of Dr. S. G. Weissberg of

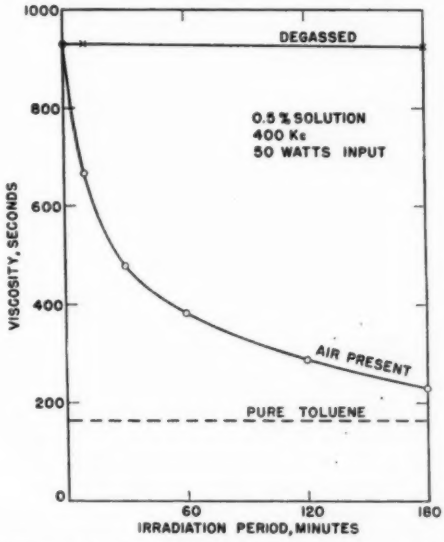


FIG. 1. Ultrasonic depolymerization of polystyrene in toluene.

[1] H. Mark, J. Acous. Soc. Am. **16**, 183 (1945).
[2] K. Sollner, Chem. Rev. **34**, 371 (1944).
[3] H. Freundlich and D. W. Gillings, Trans. Faraday Soc. **34**, 649 (1938).
[4] By "cavitation" we mean the formation and violent collapse of small bubbles or voids in the liquid, as a result of pressure changes which occur, for example, in a sound wave.
[5] G. Schmid and O. Rommel, Zeits. f. physik. Chemie **185A**, 97 (1939).
[6] Prepared by polymerization at room temperature for several months.

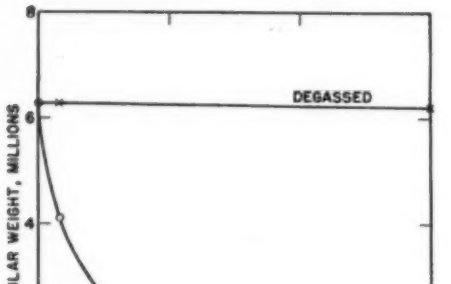


FIG. 2. Ultrasonic depolymerization of polystyrene in toluene.

the National Bureau of Standards, was dissolved in toluene to give a 1 percent solution. A 50 ml sample of this solution, containing dissolved air, was irradiated[7] for 10 minutes (at a power level of 50 watts d.c. in the plate circuit of the output tube of the ultrasonorator) and then diluted to 100 ml. After this treatment the viscosity (measured at 30°C with an Ostwald-Fenske viscometer) was 665.7 seconds, compared to 929.3 seconds for an unirradiated diluted solution and 163.6 seconds for the pure solvent toluene. But when a similar sample was degassed first by boiling under vacuum, then cooled in the absence of air, irradiated, and diluted to 100 ml, no viscosity change was found (Fig. 1). Also, there were no cavitation bubbles in the degassed solution, although many were visible in the solution containing air.

Similarly, 3 hours irradiation gave only a very slight change (to 925.7 seconds) in a degassed solution, but a large viscosity decrease (to 270.9 seconds) in the solution containing air. Several intermediate points are also shown on Fig. 1. These viscosity decreases were permanent; no change could be detected after the solutions had stood for 20 hours.

The extent of depolymerization corresponding to the observed viscosity decrease was calculated from the equation

$$[\eta]=KM^a,$$

where M is the molecular weight, and $K=1.6\times10^{-4}$ and $a=0.66$ are two constants (obtained by extrapolation from data in the literature[8]) characteristic of toluene solutions of polystyrene which has been prepared

[7] The container was a 50×400 mm test tube, closed at the top with a ground joint through which were sealed two gas-delivery tubes, one extending almost to the bottom of the container. This made it possible to change the nature of the dissolved gas in the solution. The ultrasonic waves used in these experiments had a frequency of 400 kilocycles.

[8] H. Mark in *Physical Methods of Organic Chemistry*, edited by A. Weissberger (Interscience Publishers, Inc., New York, 1945), p. 146; certain errors in this table were corrected for us by Dr. S. G. Weissberg.

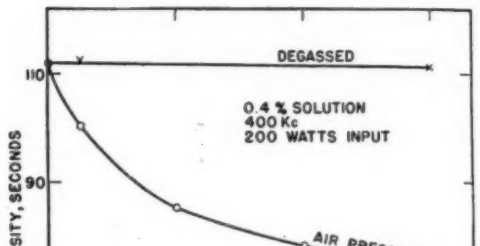


FIG. 3. Ultrasonic depolymerization of hydroxyethyl cellulose in water.

by polymerization at 30°C, as was the present sample. The intrinsic viscosity $[\eta]$ is defined in terms of the viscosity of pure solvent η_0 and the viscosity η_c of a solution of concentration c:

$$\eta=\lim_{c\to0}\frac{\eta_c-\eta_0}{c\eta_0}=\lim_{c\to0}\frac{\eta_{sp}}{c}.$$

By means of successive dilutions, η_{sp} was evaluated for each irradiated solution at concentrations of 0.50, 0.20, and 0.10 gram/100 ml, and the results extrapolated to infinite dilution to give $[\eta]$. Figure 2 shows that the viscosity-average molecular weight decreases from about 6,000,000 to 600,000 in the ordinary solution, but in the degassed solution the decrease is negligible—of the same order of magnitude as the experimental error.

This result indicates that cavitation is responsible for the depolymerization. In order to determine whether the same is true in water solutions, analogous experiments were performed on a 0.80 percent aqueous solution of hydroxyethyl cellulose WS-1000 (obtained through the courtesy of the Carbide and Carbon Chemicals Corporation). Here again, ultrasonic irradiation

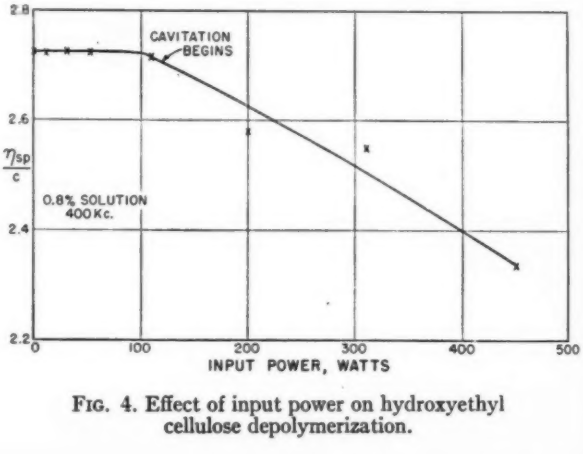


FIG. 4. Effect of input power on hydroxyethyl cellulose depolymerization.

for 3 hours at 200 watts input—followed by dilution to 0.4 percent—gave a very slight viscosity change (from 112.2 to 111.3 seconds) in degassed solutions, but a large decrease (to 73.6 seconds) in ordinary solutions; distilled water had a viscosity of 61.1 seconds in this viscometer. This is represented graphically in Fig. 3, together with several intermediate points.

An even more direct relation between cavitation and depolymerization is demonstrated in Fig. 4. It was previously observed that no cavitation bubbles could be detected in water irradiated at the lowest power levels, even with intense transverse illumination. However, bubbles could always be found at input powers above a given threshold, the value of which depended on the experimental conditions; in viscous solutions the threshold was higher and less sharp. In the present work, by subjecting successive portions of hydroxyethyl cellulose solutions to 5-minute irradiations at increasing power levels, it was found that no depolymerization occurred at the low levels where cavitation was absent. But just about at the power where cavitation began, depolymerization also began, and became greater as the power was increased.

It might be supposed that the observed degradation is due to oxidants which are known to be produced by irradiation of solutions containing nitrogen or oxygen.

TABLE I. Effect of dissolved gas on ultrasonic depolymerization.

Dissolved gas	Viscosity of 0.20% solution of polystyrene in toluene, after 30-minute irradiation	Viscosity of 0.40% solution of hydroxyethyl cellulose in water, after 15-minute irradiation
None	256.6 seconds	112.7 seconds
Air	177.6	100.3
Helium	175.0	100.0
Carbon dioxide	184.4	112.3

This view is invalidated, however, by the data in Table I; substantially the same amount of depolymerization occurs even when purified helium is the only gas present. Carbon dioxide also serves just as well in toluene, but not in water solution.

CONCLUSION

It is known that cavitation does not occur at moderate negative pressures in solutions which have been freed of dissolved gas. All concomitants of the ultrasonic wave, other than cavitation, are present in degassed solutions, but no depolymerization takes place. Therefore, it is cavitation which is responsible for the depolymerization caused by ultrasonic irradiation, in the polystyrene-toluene and hydroxyethyl cellulose-water systems.

A Three-Stage Electron Microscope with Stereographic Dark Field, and Electron Diffraction Capabilities

M. E. HAINE,* R. S. PAGE, AND R. G. GARFITT
Metropolitan Vickers Electrical Company, Ltd., Manchester, England
(Received August 2, 1949)

An electron microscope is described in which a range of magnifications between 1000 and 100,000 times can be covered continuously. This is achieved through the use of an intermediate projector lens.

The principles of design are briefly discussed and the construction of the microscope and the high frequency-operated 100-kv power unit is dealt with in some detail. The instrument is provided with objective apertures to enhance image contrast and to permit operation with dark field illumination.

Data is given relating to the electron optical system and it is shown how the instrument can be used to obtain electron diffraction patterns characteristic of small areas of a specimen under microscopic observation. The alignment procedure is briefly described and some typical micrographs are shown.

I. INTRODUCTION

FOLLOWING the work on the design and construction of a two-stage electron microscope Type E.M.2,[1] it was decided to undertake the development of a new instrument, the design of which would aim at obtaining the greatest possible simplicity and economy compatible with the best performance.

This paper describes the design principles evolved for the new instrument (Fig. 1) and explains the versatility in operation and also the mechanical, electrical, and electron optical details.

Previous electron microscopes have used a two-lens system, except that built by Marton[2] which used a three-lens system. The relative advantages of the two systems were considered and it soon became apparent that a three-lens system offered many advantages in providing a more compact design and a much greater versatility. Three-stage instruments have been de-

* Now with the Associated Electrical Industries Research Laboratory, Aldermaston, Berks, England.

[1] M. E. Haine, "The design and construction of a new electron microscope," J.I.E.E. 94, Part 1, No. 82 (October, 1947).

[2] L. Marton, "A 100-kv electron microscope," J. App. Phys. 16, 131 (1945).

veloped also coincidently with our own by le Poole[3] and Liebmann.[4]

A two-stage instrument consists of lenses separated by distances of three to four times the length of a lens. This means a long and hence not very rigid column. Ideally a column consisting of lenses, and space enough to allow for the necessary aligning mechanisms, would suffice. If we fix the ideal length per stage at a value l_0 and assume the lenses all have focal lengths f_0, then with n stages the magnification is given approximately by

$$M=(l_0/f_0)^n.$$

Thus,

$$n=\log M/\log(l_0/f_0).$$

For a maximum magnification of 100,000 times, $l_0=20$ cm, and $f=0.4$ cm, n becomes 3 to the nearest integer. The resulting three-stage column is 60 cm long compared with 250 cm for the corresponding two-stage instrument.

Further consideration shows that a much wider range of magnification for a given specimen setting is possible. For instance, whereas in a two-stage instrument the magnification can only be varied over a range of about five to one without changing the specimen position, in the three-stage instrument a range of nearly one hundred to one is possible. This further simplifies design as it obviates the necessity for an intermediate viewing screen for aligning purposes.

A further advantage results by virtue of the ease with which diffraction photographs can be obtained by imaging the diffraction pattern at the back focal plane of the objective lens on to the object plane of the second projector lens, using the first projector lens at unity magnification. The detailed mechanism involved is described in later paragraphs.

In addition to using a three-lens magnifying system, it was thought possible to reduce the over-all volume of the apparatus sufficiently to keep the "pump-down time" low enough to allow complete evacuation in two or three minutes and thus obviate air locks for specimen and photographic plate. This was considered particularly important in view of the difficulties already present in the mechanical stage design which are much aggravated if an air lock is necessary.

Our own past experience and that reported to us from other workers in this field made us decide to use an electron accelerating voltage of 100 kv. There seems little doubt that this represents an appreciable improvement on the 50 kv previously used, and yet it is evident that to increase the voltage to much over 100 kv involves considerable complication and expense.

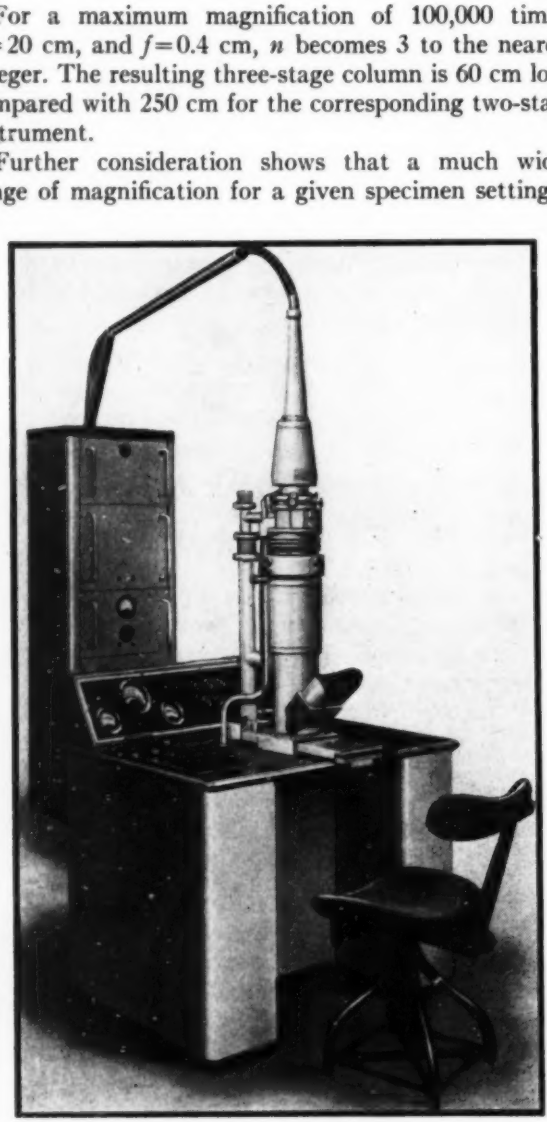


FIG. 1. Photograph of the E.M.3 electron microscope.

II. DESCRIPTION OF MICROSCOPE COLUMN

The microscope column (Fig. 2) is built up in eight sections: The electron gun, condenser lens, specimen chamber, objective lens, intermediate alignment section, twin projector lenses, viewing chamber, and camera. Each unit is supported directly by the section underneath and vacuum sealed by rubber gaskets. Mechanical rigidity, a first consideration in electron microscope design, is maintained by the use of the type of rubber gasket seal used on the Type E.M.2 instrument which allows the opposing flanges to be clamped together in metal to metal contact.

A. The Electron Gun and Condenser Lens

The electron gun is totally enclosed, utilizing a screened cable entry bushing insulator (103) to conduct the high voltage lead (101) to the cathode. Thus no high voltage is exposed at the electron gun and the necessity for a cumbersome screening compartment is obviated.

The three-core high voltage cable is terminated in a three-contact plug assembly which fits into the central bore of the conical porcelain insulator and makes contact with the cathode assembly of the gun. A conical screen surrounds the whole insulator and serves to connect the cable screening direct to the main body of the microscope. The cathode assembly (111) is of more

[3] J. B. le Poole, "A new electron microscope with continuously variable magnification," Philips Tech. Rev. 9, No. 2 (1947).
[4] G. Liebmann, "A new experimental electron microscope," J. Sci. Inst. 25, No. 2 (February, 1948).

or less conventional design, consisting of a tungsten hairpin filament surrounded by a biased cathode shield. The anode (117) has a central aperture to allow the beam to pass out of the gun and also a concentric ring of holes for pumping. The gun is surrounded by a lead-lined x-ray shield (118).

The electron gun is mounted directly on top of the condenser lens; no transverse alignment of the lens with respect to the gun axis has been found necessary.

The electron gun and condenser lens assembly is mounted on an aligning mechanism which allows it to be translated laterally or tilted about a point in the object plane for electron optical alignment. The mechanism employed is similar to that used in the Type E.M.2 instrument, but it has been reduced in size and made more rigid. The lateral movement is obtained by sliding two flat surfaces over one another with the adjusting screws (211) while the tilt is obtained by sliding two spherical surfaces centered at the specimen by means of adjusting the screws (212). Vacuum sealing is effected by a flexible metal bellows. The complete unit is mounted on top of the object chamber and sealed by a rubber gasket.

B. The Specimen Chamber and Mechanical Stage

The specimen chamber (Fig. 3) consists of a brass cylinder brazed to upper and lower sealing plates (302, 303). To provide access to the mechanical stage the front section of the cylinder is cut away to accommodate the door (304), which is sealed with a flat rubber gasket held to the door by an internal retaining plate. A swing catch pivoted at the center of the door engages in slots in the upper and lower flanges. Atmospheric pressure is sufficient to seal the door without addition of any auxiliary pressure, and the catch is provided merely to hold the door closed before evacuation commences. A pumping manifold connection is made at the back of the object chamber.

The object chamber is fixed directly on top of the objective lens by flange bolts, and sealed with a rubber cord gasket.

The object stage itself consists of a cup-shaped carriage (306) which slides directly on the upper face of the objective lens. These sliding surfaces are polished to reduce the friction so that a smooth movement is obtained, while two spring-loaded plungers hold the surfaces in close contact.

The specimen holder (309) is locked into the stereo block (310) by means of the finger-controlled screw (311), and is shown in detail in Fig. 4. The holder can be extracted from the stage through the chamber door with the aid of the special tweezers (Fig. 4) which engage with an internal groove in the holder.

The stereo block into which the specimen holder fits is mounted on two pointed pegs which rest in grooves in the base of the carriage, lying approximately in the plane of the object. The upper face of the block is cylindrically machined about an axis containing the

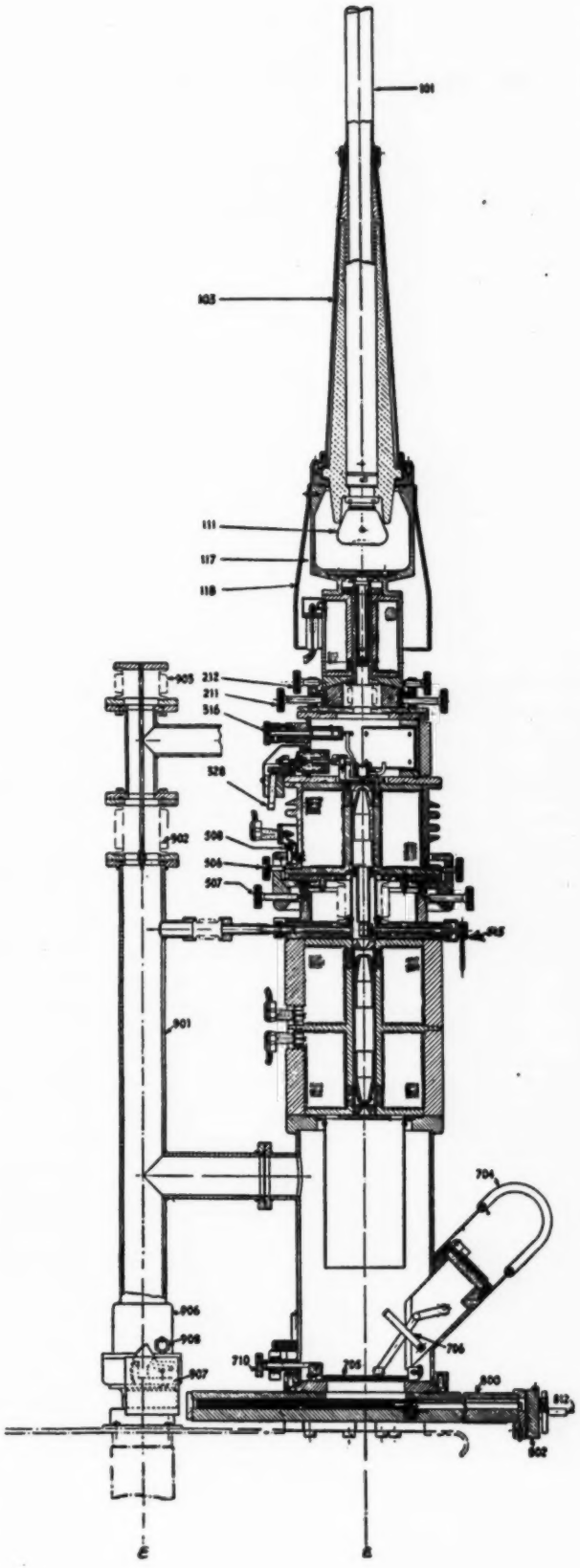


FIG. 2. Cross-sectional drawing of the microscope column.

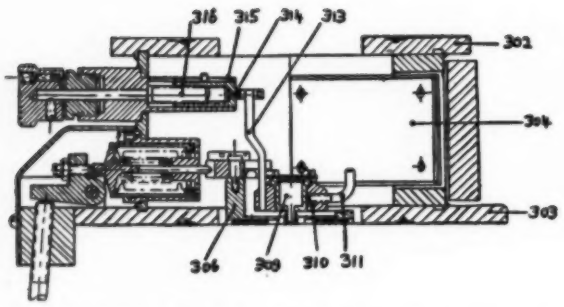


FIG. 3. Cross-sectional drawing of the specimen chamber.

points of the pegs. Leaf springs fastened to the carriage bear on this cylindrical surface, allowing the block to be tilted through a total angle of 10° about the axis in the object plane. Tilting is effected by a vertical lever (313) attached to the stereo block and engaging the tilting mechanism at the top of the specimen chamber. The tilting control consists of a lead screw (316), entering the vacuum through a rubber gland seal, which moves the block (314), keyed to the fixed tube (315), laterally. A rectangular opening in the block allows clearance around the lever so as not to interfere with the mechanical stage movement.

Motion in two directions at right angles is transmitted to the object stage through a lead screw and bell crank system from two knobs at hand level. Careful attention to the friction, controlling springs and rigidity of the operating mechanism gives smooth control to 0.1μ or better. The relatively small mass of the stage and close contact with the objective lens ensures thermal equilibrium between these being reached in a short time.

C. The Objective Lens and Intermediate Aligning Mechanism

The objective lens is of conventional design; the bore is made vacuum tight by a brazed copper spacer in the gap and is ground to give accurate cylindricity.

The pole pieces are asymmetric about the lens center, in order to bring the object focal plane to a position just above the top of the lens. A special grinding technique has been developed to ensure a minimum degree of eccentricity and ellipticity of the pole piece bores to minimize astigmatism. Further improvement in this direction is achieved by making the bore diameter as large as practicable with no increase in focal length, as by so doing the effect of residual mechanical and magnetic defects as well as the primary spherical aberration is reduced.

A focal length of 4 mm is obtained at 100 kv with the specimen plane slightly above the level of the upper lens face and a pole piece diameter of 2.5 mm. The pole pieces are finished by precision grinding to be an accurate push fit in the bore of the lens and can be removed for cleaning relatively easily.

The alignment section immediately below the objective lens provides two movements, one for translating (506) and one for tilting (507) the objective lens and illuminating system with respect to the projector lenses. The tilt takes place about a point in the object plane of the second projector lens. Each movement is controlled by a set of four self-locking screws.

D. Objective Apertures

The microscope has been fitted with two types of objective aperture; one being a relatively large aperture intended to give an increase in image contrast and to be easily inserted and removed. It takes the form of a hollow phosphor bronze cylinder closed at one end by a thin wall into which is drilled the aperture hole 100μ in diameter. The cylinder fits closely into the bore of the objective lens and can be inserted and removed by means of a special tool through the specimen chamber door. The operation takes slightly longer than that of changing the specimen. The plane of the aperture is located approximately at the back focus of the lens so that electron diffraction patterns cannot be obtained with the aperture in position.

The alternative type of aperture takes the form of an assembly which fits into the lower pole piece of the objective lens from the underside. The aperture, again located at the back focus of the objective lens, can be preset in lateral position by four grub screws and is centered before assembly of the objective lens by means of a low power microscope and "V"-block. This precentering allows the use of a smaller aperture. The assembly is thus semipermanent and precludes the use of electron diffraction or low magnification dark field imaging. Normal dark field images can be obtained, however, by tilting the illuminating system sufficiently for the direct beam to be intercepted by the aperture.

In addition, a special intermediate section has been developed for the instrument which carries an adjustable objective aperture assembly, details of which can be seen in the cross section of Fig. 5. The aperture is carried on the detachable end of a tube (551) which slides vertically inside a carrier tube (552) supported in a spherical seating (553) about 3 cm below the center of the objective lens. Thus, transverse movement of the lower end of the carrier tube will translate the aperture in the lens field. A main supporting tube (554) carries the spherical seating at the upper end and a cylindrical block at the lower end. Two fine-threaded

FIG. 4. The mechanical stage, specimen holder, and insertion tweezers.

screws through this cylinder allow adjustment of the position of the carrier tube. These screws extend to knobs outside the walls of the alignment section.

Vertical movement of the aperture tube is controlled by the fork (556) engaging with two pins in the lower end of the tube. The fork is fastened to a shaft projecting through the walls, and a knob on the end engages with stops suitably placed on the external wall.

The normal operating position for the aperture is about 4 mm below the lens center. When it is not required it can be raised to just below the level of the specimen plane so that the cone of electrons leaving the object area is unrestricted. When the specimen is removed, the aperture can be lifted even further so that it projects above the level of the lens face. This enables the cap containing the aperture disk itself to be removed or exchanged without having to dismantle the whole lens assembly.

The aperture used is a thin copper disk produced by Kodak. The hole sizes so far used are 30 and 60μ in diameter.

The lower plate of the aligning section carries a pumping port connected to the vacuum manifold, also another movable aperture system (515) inserted at the level of the first intermediate image. This is used in obtaining electron diffraction patterns as described in Section III. It consists of a cylindrical cap on the end of a shaft controlled by a knob. Two clearance holes in the cap allow the beam to pass unrestricted for normal operation while rotation of the knob through 60° or 120° brings two small apertures of 50μ or 250μ diameter, respectively, in line with the axis of the beam. A screw backstop is provided for initial alignment of these apertures.

E. The Double Projector Lens, Viewing Chamber, and Camera

The construction of the double projector lens is clearly shown in Fig. 2.

The yoke, copper spacers, and end plates are brazed together, giving a vacuum-tight bore. The lens winding is wound directly on the double bobbin so formed and the two halves of the outer iron shroud slide over the whole assembly. The bore is precision ground and sufficient accuracy is obtained to render individual alignment of the two lenses unnecessary.

The viewing chamber is provided with a single wide viewing port with a lead glass window and eyeshield (704). On the sides of the viewing port are fitted the controls for raising the fluorescent screen (705) and for positioning a magnifying lens (706) to aid correct focusing.

A pumping port connects the viewing chamber to the vacuum manifold and the top flange supports a cylindrical μ-metal screen. The viewing chamber is fastened on blocks mounted directly on the main cross members of the desk frame and the camera is strapped underneath

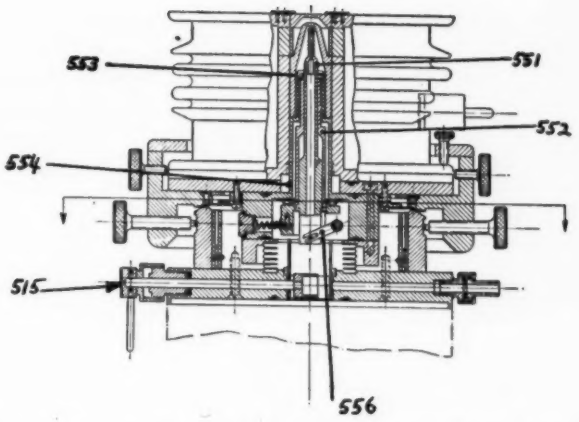


FIG. 5. Cross-sectional drawing of the intermediate alignment section with the adjustable objective aperture.

the lower flange, so that it can be removed without disturbing the rest of the microscope.

The camera consists of a shallow rectangular box with a transverse shaft operating the cassette tray through a pinion and rack. The door (802) is sealed by a flat rubber gasket and held in place by a spring clip. The drive for the racking pinion is brought out through a gland seal and bevel gears to knob (812) at the front right-hand side of the camera and a simple position indicating mechanism is incorporated.

The camera will take two $3\frac{1}{4}\times3\frac{1}{4}$-in. plates end to end, on which two, four, or eight micrographs can be taken. The frame size is selected by masks attached to the fluorescent screen; the appropriate mask is selected by the knob (710) at the rear of the viewing chamber.

F. Vacuum Equipment and Desk

The pumping equipment consists of a Metro-Vick Type D.R.I. two-stage rotary pump backing a Type 02 oil diffusion pump. The pumping system is manually controlled and provided with simple interlocks to prevent damage due to incorrect operation. The backing pump can be connected direct to the microscope column for rough pumping or to the diffusion pump for normal operation.

A simple fine side vacuum valve is incorporated between the diffusion pump and pumping manifold so that the microscope can be open to the atmosphere without shutting down the pumps. The over-all pumping time from atmospheric pressure is about two minutes.

The pumping manifold leading from the diffusion pump via the fine side valve has three connections to the microscope column (Fig. 2) as previously described; one at the viewing chamber, a second at the intermediate section, and a third at the object chamber. Flexible connections are made in the latter two to allow for the necessary alignment of the microscope column. The bending moment exerted on the column by air pressure on the top flexible connection (902) is com-

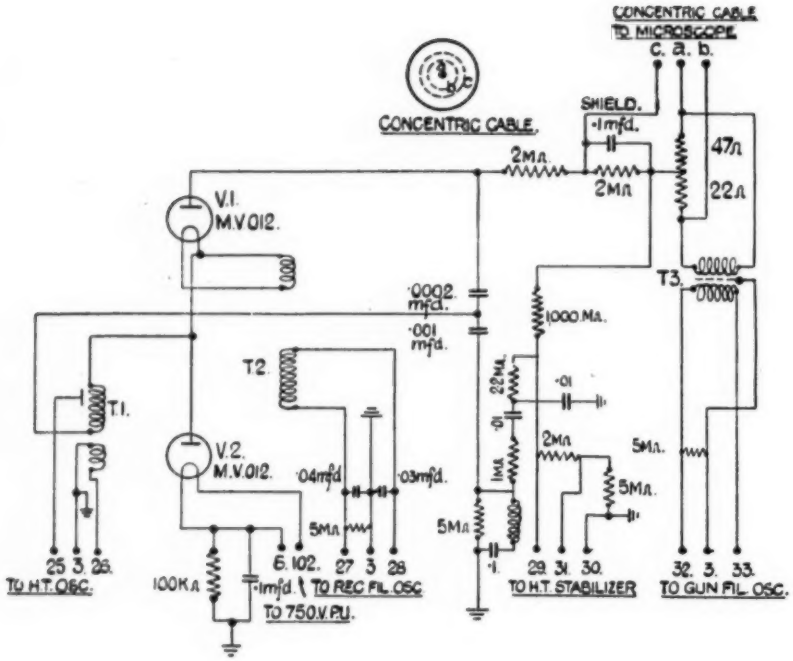


FIG. 6. Circuit diagram of the oil tank assembly.

pensated by a further flexible bellows (903) on the top of the manifold. The top plate sealing this bellows is mechanically connected to a spider in the lower end of the connecting bellows by a stay rod. Thus no resultant moment acts on the column.

The microscope desk (Fig. 1) is constructed from an angle iron framework with sheet metal paneling, and is mounted on rubber-bonded feet to insulate against vibration. A sloping panel at the rear of the desk displays beam current, lens current, and vacuum gauge meters, indicating lights and the a.c. mains circuit switches and fuses. The microscope controls are all grouped on a hinged panel just to the left of the column, while the lens current stabilizers are built on subpanels fastened to the control panel, providing easy access for maintenance.

All internal wiring takes the form of multiple cables or looms interconnecting each unit, terminated with multiple terminal blocks and spade end-coded leads.

G. The High Voltage Power Unit

This follows what may now be regarded as conventional practice for this type of instrument, the high voltage being obtained by rectification of the output of a resonant high frequency transformer fed from a valve oscillator.[5] Stabilization is obtained through negative feedback from a high resistance potential divider connected across the output circuit, via a d.c. amplifier, to the screen circuit of the oscillator valves, four 45-volt batteries providing the stable comparison voltage.

The complete unit contains three separate oscillators supplying power to the high voltage circuit, the high potential rectifier valve filament, and the electron gun filament. The operating frequencies are, respectively, 30, 50, and 50 kc/sec. All the oscillator and stabilizer components are grouped on a single rack-mounted chassis, while a second chassis contains the low voltage power supplies to the oscillators. The high voltage unit, which employs a Greinacher voltage doubler circuit, is housed in a brass oil-filled tank, and the electron gun supply is fed out through a shielded concentric cable. The voltage output of the set is varied by selection of the number of reference batteries in circuit, four 25-kv steps being provided. On the 25-, 50-, and 75-kv ranges, maximum output current is about 1 ma. and 0.4 ma. on the 100-kv range.

The electron gun filament power is remotely controlled from the microscope desk by variation of the anode supply voltage to the appropriate oscillator. Switching of the high voltage supply is accomplished by means of push buttons and relays, which also afford automatic overload protection. Figures 6 and 7 are circuit diagrams, respectively, of the high voltage and oscillator-stabilizer units.

III. ELECTRON OPTICAL PERFORMANCE

A. Illuminating System

The triode electron gun operates with an automatic bias derived from a variable resistor connected between the cathode and the cathode shield. For a given setting of the bias resistor there is thus a value of filament heating current which gives the correct lens action at the cathode shield aperture and produces an intensity maximum at the beam cross-over situated near the anode aperture.

[5] O. H. Schade, "Radiofrequency-operated high voltage supplies for cathode-ray tubes," Proc. I.R.E. 31, No. 4 (April, 1943).

Operating a triode gun under these conditions minimizes the effect of bad alignment of the filament in the shield aperture and also gives maximum brightness conditions. An image of the beam cross-over at approximately unit magnification can be formed in the object plane of the objective lens by the magnetic condenser lens, so that by operating the microscope at low power an image of the crossover can be projected on to the viewing screen. The cross-sectional diameter of the crossover determined in this way has a minimum value of about 0.005 cm, the actual value depending on bias setting and on the height of the filament in the shield. Under optimum conditions, the beam divergence from the crossover is approximately 0.03 radian. With a total beam current of 200 μa, the current passing through the condenser aperture, which subtends an angle at the cathode of 4×10^{-3} radian, is

$$200\times10^{-6}\times\left(\frac{4\times10^{-3}}{3\times10^{-2}}\right)^2=3.56\ \mu\text{a}.$$

The mean current density at the object is therefore

$$\frac{3.56\times10^{-6}}{\pi(2.5\times10^{-3})^2}=0.18\ \text{amp./cm}^2.$$

For adequate brightness on the final screen, a current density of 10^{-11} amp./cm^2 is required; hence the gun gives a sufficient intensity for operation at a magnification of $(0.18/10^{-11})^{\frac{1}{2}}=135{,}000$ times at critical illumination, or at an angular aperture of 5×10^{-4} radian, the visibly usable magnification is

$$1.35\times10^5\times\left(\frac{5\times10^{-4}}{4\times10^{-3}}\right)^2=2100\ \text{times}.$$

B. Lens System

Consideration of minimum aberration in the objective lens calls for the use of the shortest possible value of focal length. For a given pole piece design the minimum focal length is determined by magnetic saturation of the iron. Scaling up the dimensions of the pole pieces while maintaining constant field strength in the gap again shortens the focal length. Limitations here arise due to excessive increase in the excitation ampere turns required and also due to the fact that the lens field becomes so extensive that the back focus comes within the field, after which rays are deflected back toward the axis, and hence the focal length no longer decreases. A further advantage in scaling up the pole pieces results from the relative reduction in the astigmatism resulting from mechanical inaccuracies.

A further limitation in the focal length arises because of the difficulty of accommodating the object holder in

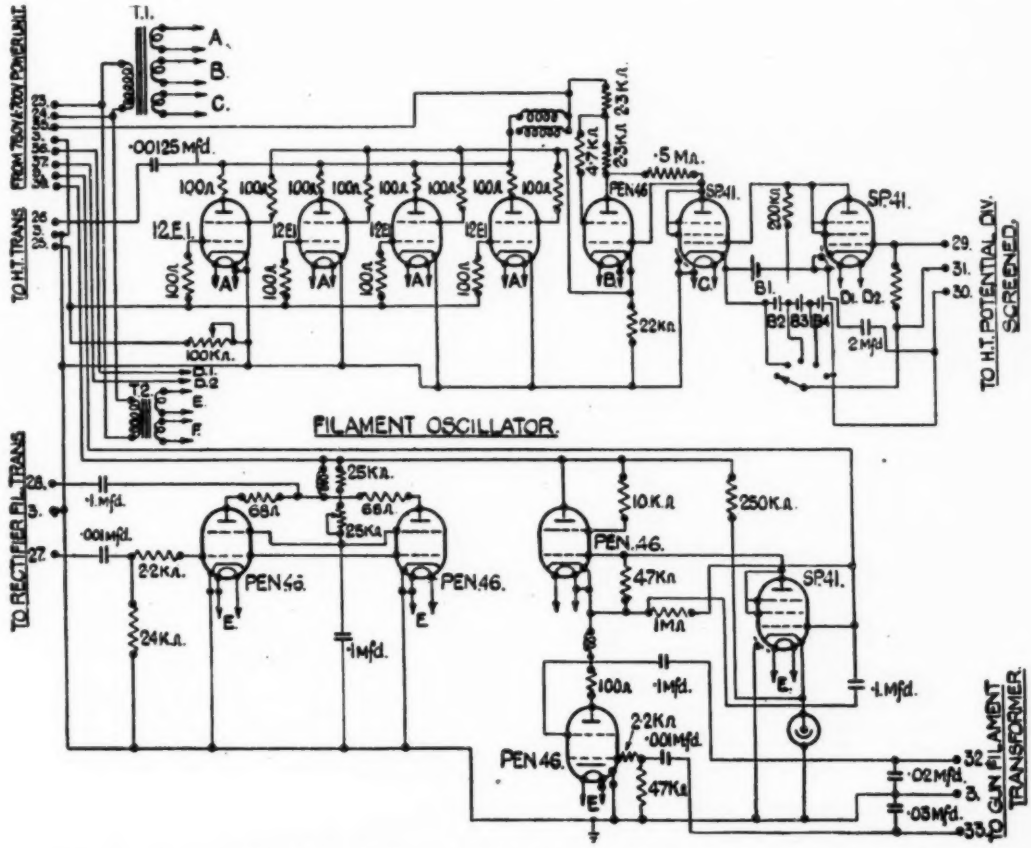


FIG. 7. Circuit diagram of the H.T. oscillator and stabilizer.

TABLE I. Magnification ranges.

Accelerating voltage	Low magnification undistorted	High magnification
100 kv	600–2000	2000– 60,000
75 kv	1000–3000	3000–100,000
50 kv	1500–5000	5000–100,000+

a position very close to the lens without severely restricting its movement. For this reason a focal length of 4 mm is used, and the pole piece aperture chosen to be as large as possible for saturation at the maximum excitation. The maximum objective magnification is then about 35 times.

The three-lens system with the adjustable aperture provides a very flexible imaging system which can operate under three distinct combinations of focal lengths. These combinations provide for (1) high range magnification, (2) low range magnification, and (3) electron diffraction.

C. Magnification Range

The ranges of magnifications obtainable are shown in Table I. On the high magnification range the two projector lenses are used in the normal way, that is, both producing inverted magnified images. The ray path for this condition is shown in Fig. 8(A). To avoid excessive pincushion distortion† at the lower magnifications, the second projector is operated at high power and gives most of the magnification (about 50× at 100 kv). For an 8-cm diameter image, the second projector object field is then 0.16 cm in diameter, so that the bore of this lens can be made small without introducing distortion. Thus, for an over-all magnification of 2000 and first-stage gain of 35, the second lens must give unity magnification. The higher magnification can be obtained now by increasing the first projector power up to a maximum when it works at a focal length of about 4 mm, giving a stage magnification of 35× at 100 kv. This corresponds to an over-all magnification of 60,000×.

To obtain the low magnification condition, the first projector is used to form a reduced erect image by operating at low power. The ray path is shown in Fig. 8(B). The extent of the imaging beams at the projector lens aperture is now of the order of 0.3 cm so that the bore of the lens must be double that required by the high magnification condition in order to avoid barrel distortion. The wider bore is adopted though it slightly reduces the highest magnification obtainable.

D. Electron Diffraction

In this instrument an electron diffraction pattern corresponding to a small selected area of the specimen under examination can be formed immediately and without disturbing the specimen. This is achieved by using the first projector lens to form an image of the diffraction pattern existing in the back focal plane of the objective lens at approximately unit magnification in the object plane of the second projector lens. This image is then enlarged on to the final screen, or photographic plate. This method of obtaining diffraction patterns was originally used by Boersch.‡[6]

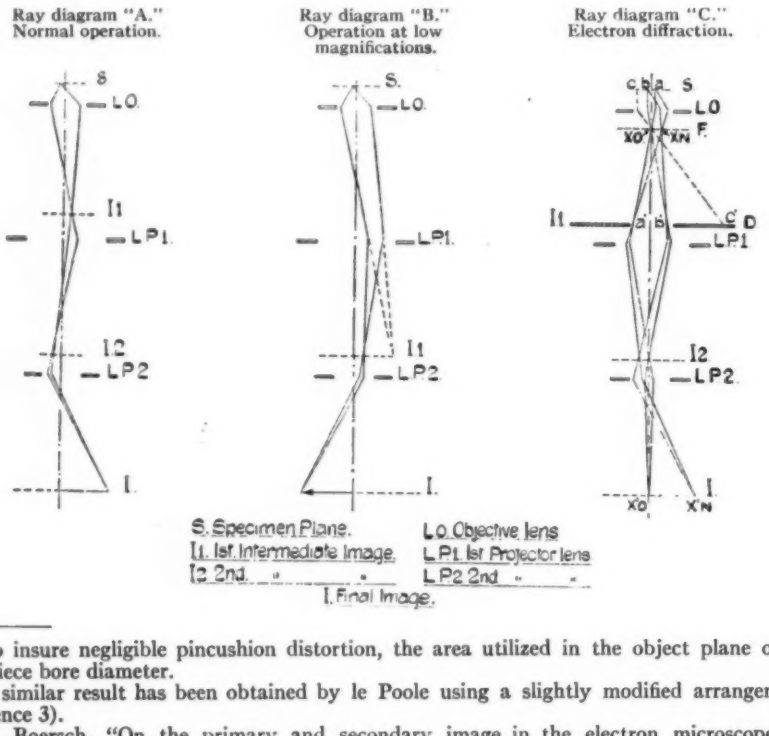


FIG. 8. Diagram of the ray paths.

† To insure negligible pincushion distortion, the area utilized in the object plane of the projector should be one-fourth of the pole piece bore diameter.

‡ A similar result has been obtained by le Poole using a slightly modified arrangement incorporating an additional weak lens (reference 3).

[6] H. Boersch, "On the primary and secondary image in the electron microscope II. Structure investigation by means of electron diffraction," Ann. d. Physik. 27, 75 (1936).

Under these conditions, the objective lens is used in the normal manner and therefore produces an intermediate magnified image of the object just in front of the first projector lens. A stop placed here will limit the field of the object observed in normal microscopic observation and will also, therefore, in the diffraction usage, limit the area of the object from which diffracted electrons are transmitted. The stop can be relatively large in diameter since the effective diameter is reduced by the magnification of the objective lens (35×). Thus, in practice, two stops 50 and 250μ in diameter are used, corresponding to fields of 1.5 and 7μ at the object. The ray path is shown at Fig. 8(C).

The maximum angle of diffraction allowed for is 0.1 radian. If the whole of the beam contributes to the intermediate image, the confusion disk diameter, given by $C_S\alpha^3M\cdot(C_S=0.5$ cm) becomes 175μ which is larger than the smaller stop. The effect of this when using the stop depends on whether the paraxial or marginal electron rays are focused at the aperture. Most of the intensity will normally be in the paraxial beam and if this is focused, then some of the energy from the outer rings will be lost and some energy from the outer rings of any crystals outside but near the defined area will appear. A better condition is for the intermediate image to be underfocused by such an amount as to allow all the outer ring electrons to pass through the aperture.

The diffraction pattern rings in the back focal plane of the objective will have a diameter given by

$$D=2\alpha_D f,$$

where α_D is the angle of diffraction and f the focal length of the objective lens. Toward the outside of the pattern some distortion is introduced by spherical aberration. This has the effect of reducing the diameter of the outer rings by an amount $C_S\alpha_D^3$. For a maximum angle of diffraction of 0.1 radian and a focal length of 0.4 cm, the ring diameter becomes $D=2\times0.1\times0.4$ cm $=800\mu$ and the distortion is $\Delta D=0.5\times(0.1)^3$ cm $=5\mu$ or a little under one percent. A simple correction can be applied for this error.

The final projector lens provides sufficient magnification to enlarge the diffraction pattern to a convenient size which is just too small for distortion in this lens to be significant.

The diffraction rings are made up of overlapping images of the electron source and therefore to give finely resolved rings the source must be small. If we operate the instrument under conditions of critical illumination we should expect to get no diffraction at all, as incoherent radiation is irradiating the object. In fact, a fair degree of coherence results from the imperfections in the cathode and condenser lenses, and even if this were not so, the Airy disk produced by the condenser lens aperture would give sufficient coherence for some pattern to result; the resolution would, however, be very poor.

If the condenser lens is switched off the electron gun virtual cathode is the source and this is demagnified about 50× by the objective into the back focal plane giving a ring width of about 1μ. To improve upon this, the condenser lens can produce an image of the source demagnified about 10×. The image is closer to the objective and hence the objective demagnification is reduced to 20× giving an over-all demagnification of 200× or an image size of about 0.25μ. This for the outer rings corresponds to a resolution of 1 in 3000, which is adequate.

Spherical aberration of the objective lens will have a further detrimental effect on the outer rings of the diffraction pattern. The diameter of the disk of confusion in the back focal plane is given by

$$\delta_S=d(C_S\alpha_D^3)=3C_S\alpha_D^2d\alpha.$$

Here α_D is the angle of diffraction and $d\alpha$ is the angle subtended by the intermediate diffraction aperture at the diffraction pattern. For the 250μ stop, $d\alpha$ becomes 0.002 radian, and if $\alpha_D=0.1$ radian, then $\delta_S=3\times0.5\times10^{-2}\times2\times10^{-3}=0.3\mu$.

If this error is added directly to that due to the finite source size, the resolving power of the system becomes 1 in 1500 with an improvement toward the center of the pattern.

E. Electron Optical Alignment

The alignment procedure employed with this instrument is very similar to that used on the Type E.M.2 microscopes.[1,7] The system is first aligned as a two-stage

FIG. 9. Staphylococcal bacteriophage and staphylococcus, shadow cast at 4:1. (32,000 times, 75 kv.)

[7] M. E. Haine, "The electron optical system of the electron microscope," J. Sci. Inst. **24**, No. 3 (March, 1947).

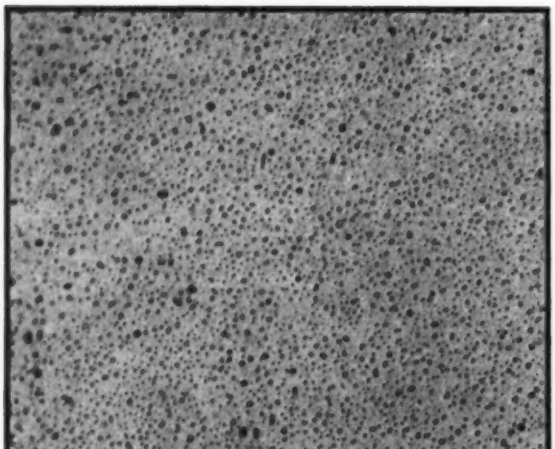


FIG. 10. Evaporated palladium gold. (72,000 times 75 kv.)

FIG. 11. Electron diffraction pattern of zinc oxide crystals, 100 kv.

microscope with the first projector lens switched off, using reversal of the objective lens current to locate the lens axis.

The first projector lens is then switched on at low power to give an image of the back focus of the objective lens in the image plane of the final projector (electron diffraction condition). Under these conditions the spot on the final screen will be displaced from the center if the first projector lens is not aligned with the axis of the objective lens. This is corrected by use of the intermediate section tilt. As this tilt is arranged to occur about the center of the second projector lens, the alignment of this lens and the objective is not upset. However, if the tilt required is large it may be necessary to repeat the initial adjustment.

The magnification can then be raised to 10 or 20,000 times and final adjustment of the gun tilt carried out.

CONCLUSION

In operation, we have found this instrument to give the promised flexibility and simplicity. The absence of air locks proves no disadvantage; the specimen or photographic plates can be changed in about three minutes while such operations as changing an aperture or a filament take very little longer.

In regard to resolving power, the average value obtained during a series of tests with heavy metal particles of about 25A diameter was approximately 35A with a best value of 25A. Edge contours have been recorded with half-widths of about 15A while spacings of as low as 12A between maxima and minima in Fresnel fringe systems have been obtained. These results indicate that a resolving power of better than 50A should be obtained under normal operating conditions.

Figures 9 and 10 show typical micrographs while Fig. 11 shows the micrograph and diffraction pattern, obtained from a field 7.0μ in diameter, of zinc oxide crystals.

ACKNOWLEDGMENTS

The authors wish to express their appreciation of the help afforded them in the development of this microscope by Mr. R. W. P. Tostevin, Mr. H. M. V. Young, and Mr. H. Nuttall, also to Mr. G. A. Barnett for his collaboration on the mechanical design.

In conclusion, the authors express thanks to Mr. F. R. Perry for his encouragement in this development, to Sir Arthur P. M. Fleming, Director, and to Mr. B. G. Churcher, Manager of Research Department, Metropolitan-Vickers Electrical Company, Ltd., for permission to publish the paper.

Letters to the Editor

Comments on the Use of Latex Spheres as Size Standards in Electron Microscopy*

CHARLES H. GEROULD
Magnesium Laboratories, The Dow Chemical Company, Midland, Michigan
October 19, 1949

BACKUS and Williams[1,2] have reported the usefulness of Dow latex 580G, lot 3584, as a standard in the determination of magnification and shadow-casting angle in electron microscopy. They attributed to these latex spheres a size of 2590±25A. Since the unique value of this material became apparent, about 225 small samples have been distributed to laboratories in this country and abroad. Seventeen of these laboratories have made determinations of the average particle size and have submitted their results which have been tabulated in Table I. It is noted that these determinations include six different techniques; electron microscopy alone, combination of light and electron microscopy, light microscopy alone, interferometry with electron microscopy, light scattering and ultracentrifugation. It appears significant that thirteen of the seventeen determinations show an average size of 2588A with a range from 2520 to 2630A. The four other determinations range upward to 3200A.

It should be mentioned that more recent light scattering studies by the two laboratories listed have indicated that the procedure used is not reliable.

It has been reported[3] that the size of this latex may be affected by variations in the procedure of specimen preparation, such as degree of dilution, size of drop added to specimen, drying speed, etc. In order to check this possibility, sixteen experiments were performed in which the dilution of the 40 percent latex with water was varied between 1:250 and 1:4000, the size of the drop added to the specimen and the drying times varied to reasonable extremes, dilutions made from a drop of latex which had been allowed to nearly dry before diluting, and dilution of the settled and caked material found in the bottom of the latex bottle. The results indicated that none or, at the most, very few significant changes took place. Twelve experiments showed variation of less than 1 percent from the mean, two between 1 and 1.5 percent and two between 2 and 3 percent. These latter two experiments may be significant or they could be the result of instrumental fluctuation. However, in both of these cases, the 2 to 3 percent variation was a decrease in size as a result of the partial drying of a drop of latex before dilution, a procedure rarely encountered in actual practice.

TABLE I. Determinations of particle size of latex 580G, lot 3584.

Contributor	Affiliation	Average size in angstroms	Estimated accuracy in angstroms	Experimental procedure			
				Instruments used	Reference specimen	Type of replica	Remarks
R. C. Backus R. C. Williams	Univ. of Mich.	2590	± 25	LM–EM[b]	Latex on glass fibers	None	
G. H. Bergold	Can. Dept. of Agri.	3100	±100[a]	LM–EM	Glass fibers	None	Instrumental[c] Calibration
J. S. Boadway	Shawinigan Chem., Ltd.	2760	± 60	LM–EM	15,000 lines/in. Grating+Latex	Collodion	Replica checked using LM
W. B. Dandliker	Harvard Med. Sch.	2590		Light scattering	—	—	
J. L. Farrant A. J. Hodge	Australian Council for Sci. and Ind. Res.	2560	± 25	EM	Glass fibers	None	Diameter of fibers determined by interferometric methods
E. F. Fullam	General Electric Co.	2590	± 50	LM–EM	30,000 lines/in. Grating+Latex	Parlodion-SiO_2	Replica checked using LM
C. H. Gerould	Dow Chemical Co.	2630	± 50	EM	15,000 lines/in. Grating+Latex	Polystyrene-SiO	
G. H. Hass	Engrs. Res. and Dev. Labs.	2620	± 40	EM	30,000 lines/in. Grating+Latex	Silver-SiO	
A. C. Jenkins	Linde Air Products Co.	2590	±130	EM	30,000 lines/in. Grating+Latex	Collodion	
H. Kahler	Nat. Inst. of Health	2600	± 65	EM	20,000 lines/in. Grating	Unknown	
R. A. Kern S. F. Kern	Eli Lilly and Co.	2580	± 50	LM–EM	Diatoms	None	Instrumental[c] Calibration
M. Martin	Godfrey L. Cabot, Inc.	2610	±130	EM	15,000 lines/in.	SiO	Instrumental[c] Calibration
O. W. Richards	American Optical Co.	3210	±240	Phase Contrast Light Microscope	—	—	Bright contrast using white light. Wet sample
E. H. Rowe	B. F. Goodrich Co.	2610	± 30	LM–EM	Opening in screen	None	Instrumental[c] Calibration
G. D. Scott	Univ. of Toronto	2500–3000	± 50[a]	LM–EM	15,000 lines/in. Grating+Latex	Formvar	Replica checked using LM
D. G. Sharp	Duke Univ.	2520	± 25	Ultracentrifuge	—	—	Particle Density—1.054±0.001
R. S. Spencer	Dow Chemical Co.	2560		Light scattering	—	—	

[a] On any one mount.
[b] LM—Light Microscope; EM—Electron Microscope.
[c] Instrumental Calibration—EM calibrated for magnification with specimen mentioned. Latex size determined using this calibration.

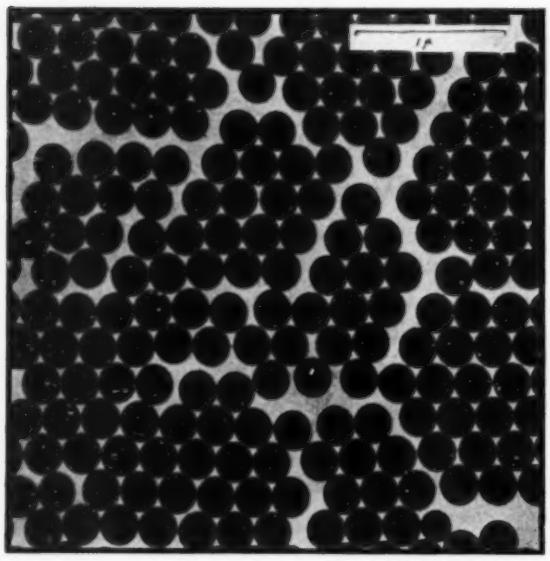


FIG. 1. Electron micrograph of single layer, latex particle array.

Since this unique batch of latex is of extremely limited quantity, the precautions in handling the material should be emphasized. Agglomeration is a danger and may be prevented by avoiding freezing, chemical contamination or change in *p*H. Agglomeration as a result of the evaporation of water from the undiluted latex is a hazard which is acute, especially in the 5 cc vials in which much of the latex has been distributed. Several laboratories have reported that their latex has completely dried even though the caps were apparently tight. It is suggested that either the bottles be periodically checked and kept filled to the top by additions of distilled water, or the whole sample be diluted 5:1 or 10:1.

The danger of bacterial contamination is not generally recognized. Experience has shown that certain bacteria can thrive and multiply in the latex resulting in eventual complete ruin of the material. This is difficult to combat, but special handling precautions and possibly refrigeration above 0°C will help.

As mentioned previously, a study was made on the effect of dilution ratio on particle size and no variation was found. However at dilutions of 1:500 and less, considerable fusing of the spheres resulted. In most cases, this fusing was slight near the wires of the supporting screen and severe in the center of the openings. An explanation may be that the surface active agents are present to such a concentration in the 1:500 or less dilutions that during drying, these agents are further concentrated to a point where they act as a plasticizer and cause fusing. It should be pointed out that the use of an atomizer may appreciably decrease this danger.

The danger of overheating in the electron microscope appears to be of importance only at the high intensities found with the biased gun. Fusing and distortion can result under such conditions. A much more real danger is found as a result of the specimen contamination which occurs under electron bombardment. These latex spheres make an ideal specimen for studying this phenomenon. Particles commonly show an increase in diameter of 4 percent to 5 percent after 20 minutes bombardment with a 300 microampere beam from a non-biased gun.

R. A. Kern and S. F. Kern† have recently discovered increases in the diameter of latex spheres as a result of shadowing, which cannot be explained by the mere thickness of the metal film on the particles. They have tentatively explained the size variations as being the result of an electrostatic charge on the unshadowed spheres which is removed by metal shadowing, providing it grounds the spheres. The possibility of such an effect should be carefully considered when these spheres are used in applications where shadowing is necessary.

Two electron micrographs are included which illustrate the particle arrays which are possible with spheres having the uniformity of diameter found in this material. Figure 1 shows a single layer particle array and Fig. 2 shows a multiple layer array in

FIG. 2. Electron micrograph of multiple layer, latex particle array. Shows areas of 1, 2, 3 and 4 layers of spheres.

which 1, 2, 3 and 4 layers of spheres can be seen. It is obvious that such arrays would be impossible without a high degree of uniformity of the spheres involved. One sphere of slightly different size will completely destroy the symmetry of the area around it.

At the present time, there are still some samples of this particular batch of latex 580G available to electron microscopists. They may be obtained by contacting the author.

* Condensed from paper given at the October, 1949, meeting of the Electron Microscope Society in Washington, D. C.
[1] R. C. Backus and R. C. Williams, J. App. Phys. **19**, 1186 (1948).
[2] R. C. Backus and R. C. Williams, J. App. Phys. **20**, 224 (1949).
[3] G. D. Scott, J. App. Phys. **20**, 417 (1949).
† To be submitted to J. App. Phys. for publication.

Erratum: Relationship between Gough-Joule Coefficients and Moduli of Vulcanized Rubbers

[J. App. Phys. **20**, 526 (1949)]

F. S. CONANT, G. L. HALL, AND G. R. THURMAN
Chemical and Physical Research Laboratories, The Firestone Tire and Rubber Company, Akron Ohio

IT has been pointed out by Dr. Sidney M. Katz of Textile Research Institute Laboratories that the thermodynamic reasoning leading to Eq. (8) of the above article is faulty, although the equation itself is valid. A correct thermodynamic derivation is possible, as Dr. Katz has demonstrated in a private communication. In view of the relatively simple mathematical derivation of Eq. (8), however, any involved thermodynamic approach appears unnecessary.

Frictional Loss in Visco-Elastic Substances

B. Gross
National Institute of Technology, Rio de Janeiro, Brazil
November 22, 1949

In two previous papers we have developed the phenomenological theory of linear visco-elastic behavior.[1] In this theory the elastic loss was associated exclusively with the relaxation mechanism which is also responsible for the frequency dependence of the dynamic modulus. On this basis the loss is uniquely determined when the dynamic modulus function is known. It seems, however, to be likely that another cause for elastic loss also exists, i.e. a "true" frictional mechanism which leads to a loss varying independently of the visco-elastic loss. The corresponding frictional force is assumed to be proportional to the rate of change of deformation and the proportionality factor, which is called frictional coefficient, is considered to be constant.[2] It might be instructive to examine experimental data for the presence of such a constant friction. The new loss component will be of importance at very high rates of deformation (or high frequencies). Therefore, the mathematical treatment now will have to take into account inertia forces too; this means that it must be based on the wave equation. But it will be sufficient to consider the unidimensional case only, because in the recent methods for the determination of the visco-elastic constants by means of sound waves the samples have the form of long thin strips which for all practical purposes can be considered as unidimensional ones.[3]

For a linear strip of material the unidimensional wave equation can be written in the form

$$q\rho\frac{\partial^2 u}{\partial t^2}dx=k, \tag{1}$$

where u is the displacement from equilibrium position of a cross section, ρ the density, and k the resultant force acting on a volume element of the sample, i.e. the difference of forces between x and $x+dx$.[4] The force now contains 3 components: (a) A component which corresponds to Hooke's law and contains the instantaneous elastic modulus E. (b) A component which takes account of the relaxation mechanism and contains the relaxation function $\bar{\varphi}(t)$. (c) A component which represents the frictional force and contains the constant frictional coefficient γ. Thus k is written

$$k=qE\frac{\partial^2 u}{\partial x^2}dx-q\int_{-\infty}^{t}\frac{\partial^2 u(x,\tau)}{\partial x^2}\bar{\varphi}(t-\tau)d\tau dx+\gamma q\frac{\partial^3 u}{\partial t\partial x^2}dx. \tag{2}$$

(1) and (2) together constitute an integro-differential equation which replaces the simpler equations of the former papers. While the general solution is involved, a particular solution of the form

$$u=u_0\exp(i\omega(t-x/c)-\alpha x) \tag{3}$$

is easily seen to exist; it represents a linear wave traveling from $x=0$ to $x=\infty$, of frequency ω, phase velocity c, and attenuation factor α. Substitution of (3) into (1) and (2) gives

$$(E-\bar{A})+i(\omega\gamma+\bar{B})=\frac{\rho\omega^2}{(\omega/c+i\alpha)^2}, \tag{4}$$

where $\bar{A}$ and $\bar{B}$ are, respectively, the storage factor and the loss factor already given in II:

$$\bar{A}(\omega)=\int_0^\infty \cos\omega\tau\bar{\varphi}(\tau)d\tau, \tag{5a}$$

$$\bar{B}(\omega)=\int_0^\infty \sin\omega\tau\bar{\varphi}(\tau)d\tau. \tag{5b}$$

Equation (4) is formally identical with Nolle's Eqs. (5). Introducing the complex elastic modulus $E^*=E_1+iE_2$ there is

$$E_1=E-\bar{A} \tag{6a}$$

$$E_2=\gamma\omega+\bar{B}. \tag{6b}$$

According to Eqs. (6), the Hookeian mechanism contributes only to the real part of E^*, the frictional force to the coefficient of the imaginary part, and the relaxation mechanism to both of them. Since $A(\infty)=0$, the instantaneous elastic modulus is obtained from a plot of E_1 *versus* frequency as the high frequency limit of E_1. $\bar{B}$ is 0 for both $\omega=0$ and $\omega=\infty$; $\gamma\omega$ is linearly increasing with frequency. Therefore, at very high frequencies where E_1 has become practically constant, E_2 should increase linearly with frequency. From the slope of the loss factor curve at high frequencies one then may read γ. The relations between $\bar{A}$ and $\bar{B}$ were already discussed in II. It will be interesting to calculate $\bar{B}$ from data of $\bar{A}$ and to see whether the result still leaves room for a component $\gamma\omega$.

[1] B. Gross, J. App. Phys. **18**, 212 (1947); **19**, 257 (1948); in the present paper quoted as I and II.
[2] A term of this type *without* any assumption about the frequency dependence has already been introduced by A. W. Nolle, J. Acous. Soc. Am. **19**, 194 (1947); J. App. Phys. **19**, 753 (1948).
[3] See R. S. Witte, B. A. Mrowca, and E. Guth, J. App. Phys. **20**, 481 (1949). D. G. Ivey, B. A. Mrowca, and E. Guth, J. App. Phys. **20**, 486 (1949).
[4] For this deduction see G. Joos, *Theoretical Physics* (G. E. Stechert, New York, 1934), p. 176.
[5] A. W. Nolle, reference 2; see A. Gemant, Trans. Faraday Soc. **31**, 1582 (1935); Physics **6**, 363 (1935).

Addendum: Interpretation of X-Ray Patterns of Cold-Worked Metal

[J. App. Phys. **20**, 885 (1949)]

B. L. Averbach and B. E. Warren
Massachusetts Institute of Technology, Cambridge, Massachusetts

In a letter under this title the next to the last sentence applied the Fourier coefficient method to the particle-size problem. "For a substance in which there is particle-size broadening only, the initial slope of the A_m *vs.* m curve gives an average particle size and the second derivative of the curve gives directly a size distribution." Unfortunately at the time of writing we were not aware that this same result had already been obtained by E. F. Bertaut[1] in France.

[1] E. F. Bertaut, Comptes Rendus **228**, 492 (February, 1949).

Bibliography on Offshore Petroleum Developments

E. N. Kemler
Southwest Research Institute, San Antonio, Texas
November 17, 1949

The Division of Oceanography and Meteorology of Southwest Research Institute recently published the first Bibliography on Offshore Petroleum Developments. Intended as a practical and useful guide, the Bibliography is a public service presentation of the Institute available to any organization, company, or individual whom it might benefit. Requests for copies should be directed to The Division of Oceanography and Meteorology, Southwest Research Institute, 312 Oil and Gas Building, Houston, Texas.

New Books

Modern Arms and Free Men

By Vannevar Bush. Pp. 273, Simon and Schuster, New York, 1949. Price $3.50.

Dr. Vannevar Bush has probably had a greater influence on *applied physics* than any other scientist of our time. His invention of the differential analyzer and its use for the solution of non-linear equations initiated the great developments in electrical calculating machines. He is famous for his leadership of the Carnegie Institution of Washington, where he has had a direct hand in the success of the 200-in. Mount Palomer telescope. He was chairman of the National Advisory Committee for Aeronautics during the critical pre-war years. But his greatest contribution is his wartime leadership of a substantial fraction of all U. S. scientists, as Director of the Office of Scientific Research and Development. In this position Dr. Bush's leadership made itself felt not only in the highest councils in Washington, but more importantly, through all levels of the vast OSRD organization, so that even the youngest men knew instinctively that all worth while developments would find effective use.

Dr. Bush gives us a very clear account of the continuing menace of the submarine, the combating of which remains the most pressing problem facing the navy of any maritime power in its mission to keep the sea lanes open for transport. In the first World War and again in the second, the German submarine held the outcome in doubt and it cost us heavily to overpower it. The technical development and tactics of sonar, radar, the sono-radio buoy, and the magnetic airborne detector, as well as greatly improved ship and aircraft-launched ordnance, finally turned the tide against the U-boat. But pro-submarine developments that came too late to re-establish the submarine's supremacy—the schnörkel, the long-range homing torpedo, and new types of fast, deep-traveling submarines are now in the hands of potential sea commerce raiders. The added fact that the submarine, along with the merchant ship, may well be the agent for delivery of the atomic bomb accounts for Dr. Bush's current insistence on vigorous anti-submarine developments.

The atomic bomb naturally occupies a central position in this book. In his considered judgment, Dr. Bush evaluates this as a weapon of extreme power but not absolute in the sense that it makes other modes of warfare obsolete. Gone, in his analysis, is the abject fear which paralyzes clear thinking and wise planning, and gone also are the overestimates which are equally fatal.

The atomic bomb presents a grave problem in its continuing strain on the peacetime economy of the nation. This is partly compensated for by the fact that of all war materials it is probably the best suited for stockpiling purposes and in addition holds potentially great possibilities for peacetime applications. In contrast to the submarine menace, the defense against the atomic bomb seems to Dr. Bush to have great promise in the future. The development of guided missiles with proximity-fused ordnance as well as the effective use of radar warning and attack systems backed up by jet fighters may well spell the doom not only of long-range atomic bombing, but also of all forms of mass bombing against an alert and fully equipped nation. That these defensive methods may be deciding factors is suggested by the experience in the Schweinfürt raid, and also in the late but very effective introduction of the proximity-fused shell for the defense of London and the battle of Antwerp. Guided missiles will thus not only extend the offensive range of artillery and the fire power of aircraft, but may even be decisive in dooming the future effectiveness of large naval ships, and in defending great cities. The use of guided missiles on intercontinental raids Dr. Bush rules out on the basis of their excessive cost as against other forms of attack.

Proper planning for a national emergency Dr. Bush emphasizes as the theme of this book. This includes a clear recognition that cost, in its broadest sense of the effective use of the manpower and industrial and scientific capacity of a nation, is the crucial element in war. He gives brief but clear examples of how the magnetron tube and early forms of radar and sonar were available to the British and ourselves but not to the Germans at the outbreak of the war. Furthermore, the fact that wartime developments in Germany failed not only to produce an atomic bomb, but more significantly that they failed to provide jet fighters in quantity for defense against bombing shows the fatal consequences of a system which ignores the power of free scientific inquiry. It is hoped that peacetime planning, and especially education, will fully realize the emphasis placed here by Bush, and elsewhere by Conant, for the necessity of proper educational opportunities for all gifted youth in the nation, regardless of their financial status. It may well be fatal to the national safety to waste their talents.

This is one of the most important books since the war. It should be required study for every person holding even a minor position in national planning, and it is to be hoped that it may be brought to the attention of many young people, some of whom may someday find themselves over night in positions of great responsibility in a national emergency.

R. S. Shankland
Case Institute of Technology